Essays on Stochastic Volatility Models with Jump Clustering

ICMA Centre

Henley Business School

Thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

Jian Chen

October 2022

Declaration of Original Authorship

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Jian Chen

Acknowledgements

This thesis would not have been completed without support from many people. Foremost, I am indebted to my supervisors Professor Andrew Urquhart and Professor Michael P. Clements for their excellent guidance, inspiring comments and encouraging suggestions through all the three years of my studies and research. They offered great helps in constructing new ideas, and provided patient support in drafting research papers. I can still remember how they guided me in doing research and build up my confidence, especially in the first year of my PhD. I benefited tremendously from the numerous discussions with them.

I also benefited from having conversations with Emese Lazar, Alfonso Dufour, Shixuan Wang, Tobias Kuna, Xiaohan Xue, Shuyuan Qi, Yu Zhang and other friends and colleagues. I feel fortunate to have had the opportunities to work with them. I also would like to thank the ICMA centre who provided financial support and the University of Reading who provided Academic Computing Cluster Service.

Additionally, I would like to express my deep gratitude to examiners Dr. Emese Lazar and Dr. Thomas Conlon for reviewing the thesis and offering many constructive comments.

Last but not the least, my special thanks go to my beloved Mengyang Xu for her support, encourages and having faith in me.

To my parents.

Abstract

This thesis investigates models of stochastic volatility which are able to accommodate the clustering of jumps typical of many high-frequency financial time series, both in terms of describing significant features of the data, and forecasting. Chapter 1 gives an overview on the jump-diffusion stochastic volatility models, clustering behaviours of jumps and contributions of this thesis to current literature.

Chapter 2 examines the clustering behaviour of price and variance jumps using high-frequency data, modelled as a marked Hawkes process embedded in a bivariate jump-diffusion model with intraday seasonal effects. We find that the jumps of both individual stocks and a broad index exhibit self-exciting behaviour. The three dimensions of the model, namely positive price jumps, negative price jumps and variance jumps, impact one another in an asymmetric manner, that is positively and significantly correlated with jump size. We estimate model parameters using Bayesian inference by Markov Chain Monte Carlo, and find that the inclusion of the jump parameters improves model fitness. We quantify the jump intensity and study characteristics of jump clusters, we find under high-frequency settings, jump clustering activities can last 2.5 to 6 hours in average, we also find that the model with marked Hawkes process models mostly outperform others in terms of reproducing two cluster-related characteristics.

Chapter 3 uses a bivariate jump-diffusion model incorporating jump clustering features by embedding a multivariate marked Hawkes process for high-frequency forecasting. In the out-of-sample period, we use a particle filter to estimate variance at each state and forward simulating return and variance distributions. We apply

a Kalman filter to correct errors that arise with microstructure noises in the high-frequency data. The simulation studies show the effectiveness of the Kalman filter. We show that the inclusion of jump clustering significantly improves the performance of high-frequency volatility forecasting and daily realised volatility forecasting. In high-frequency volatility forecasting, we find that forecasting performance is especially better with forecasting horizons of less than two hours. We also show that expected losses of two risk measures, value-at-risk and expected shortfall, can be reduced by up to 15% using models with jump clustering features.

Chapter 4 forecasts Bitcoin's returns and return jumps using a self-exciting process embedded in a stochastic volatility model. We show the existence of the jump clustering feature, which varies depending on the frequency of the data. In an out-of-sample setting, we use a particle filter to sample latent states and conduct one-step-ahead probabilistic forecasting on future jumps (underlying intensities). We assess the forecasts by a continuous ranked probability score. We further develop a statistic that takes the discrepancies between the predicted probabilities of positive and negative jumps. We find that high and low values of the difference in predicted probabilities of positive and negative jumps is able to predict returns, and that a trading strategy based on this has a Sharpe ratio of 4.36.

Lastly, Chapter 5 concludes the thesis and discusses future research.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter will introduce the background of our studies, including some concepts and recent studies on jump-diffusion stochastic volatility models and how current literature models arrivals of jumps. It will also introduce the motivations for our studies and the contributions of this thesis.

1.1 Continuous-Time Jump-Diffusion Stochastic Volatility Models

Jumps often refer to significant discontinuities in a continuous-time process. The earliest study incorporating jumps in asset prices traces back to Merton (1976). It proposes a continuous-time stochastic volatility model incorporating jumps in the price process and applies the model in option pricing. The model takes the following form:

$$dP_t = \mu dt + \sqrt{V_t} dW_t^P + \xi_t^P dN_t^P, \tag{1.1}$$

where P_t denotes the logarithm of asset prices, μ is a drift term. V_t denotes the variance, which assumes to be constant in the model. dW_t^P is a standard Wiener process. ξ_t^P denotes the size of price jump and dN_t^P is a counting process with a constant intensity λ_0. Later literature (see e.g. Heston 1993, Bates 1996) allows a

time-varying variance, which is modelled by a mean-reversion process:

$$dV_t = \gamma(\theta - V_t)dt + \sigma_V \sqrt{V_t} dW_t^V, \tag{1.2}$$

where γ and θ denotes mean reversion speed and long-run variance mean, respectively. σ_V presents volatility of volatility. dW_t^V is another standard Wiener process for the variance. A leverage effect can be captured by allowing two Wiener processes to be correlated with a correlation $\rho = corr(dW_t^Y, dW_t^V)$ (see discussions by Black 1976, Jacquier et al. 2004). In addition, Duffie et al. (2000) discovered jumps in the variance process (also see discussions by Eraker 2004), which let the variance process become:

$$dV_t = \gamma(\theta - V_t)dt + \sigma_V \sqrt{V_t} dW_t^V + \xi_t^V dN_t^V, \tag{1.3}$$

where $\xi_t^V dN_t^V$ is the variance jump component. On top of this, Duffie et al. (2000) also let two jump sizes be correlated:

$$\xi_t^P \sim \mathcal{N}(\mu_j + \rho_j \xi_t^V, \sigma_j^2), \tag{1.4}$$

where the size of price jumps ξ_t^P is assumed to follow a normal distribution with a mean being linearly correlated with the size of variance jumps ξ_t^V.

There are also many other extensions based on this model in the literature. For example, Carr & Wu (2003) introduce a stochastic volatility model incorporating Lévy jump, which allows infinite activities of jumps. Bandi & Reno (2016) add an additional price and variance co-jumps component. Also see discussions by Bakshi et al. (1997), Pan (2002), Eraker (2004), Jacod & Todorov (2009, 2010), Todorov & Tauchen (2011).

1.2 Self-Exciting Point Processes

Models introduced up till now may model jumps differently, but they make assumptions on jumps that they arrive independently with a constant underlying intensity

$p(dN_t = 1) = \lambda_0$. However, another class of stochastic point process (Hawkes process) allows the intensity to depend past jumps. It is proposed by Hawkes (1971*a*,*b*). On top of this, a multivariate version and its extension with mark values are introduced by Liniger (2009). The model not only allows impact by its past jumps (self-exciting), but also by jumps in other dimensions (cross-exciting). This section will introduce a marked Hawkes process in a general form.

We consider a counting process N_t on the σ-algebra of $\mathcal{X}$'s Borel set, $\mathcal{B}_{\mathcal{X}}$, with d dimensions $N_t = \{N_t^i\}_{i=1,\dots,d}$ observed over the time interval $t \in [0, T]$ labeled by different marks which belongs to some mark space $\kappa_i \in \mathcal{K} \subset \mathbb{R}^d$. d sequences of event times with marks are observed in pairs $\{(t_i, \kappa_i)\}$ on $\mathcal{X} \times \mathcal{K}$ and the ground process N_g is a bounded finite point process; i.e. $N_g(\cdot) = \sum_{i=1}^{d} N_i(\cdot) < \infty$. The corresponding intensity of ground processes is given by:

$$\lambda_g^i(t) = \lambda_0^i + \sum_{q=1}^{d} \vartheta_q^i \int_{[0,t)\times\mathcal{K}} \phi^i(t-s)\, \omega_q^i(\kappa_q) N_q(ds \times d\kappa) \tag{1.5}$$

where λ_0^i is defined as an immigration intensity of dimension i, which is constant; $\boldsymbol{\vartheta}$ is a branching coefficient matrix; ϕ^i is a decay function such that $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ and we assume an exponential decaying kernel $\phi^i(t-s) = \beta^i e^{-\beta^i(t-s)}, \beta^i > 0$,. $\omega_q^i(\kappa_q)$ is a impact function of marks such that $\omega : \mathbb{R} \to \mathbb{R}_+$.

Under a general marked Hawkes process, the immigrant events arrive in a Poisson process with the immigration intensity λ_0 and marks κ. They further presage arrivals of offspring events, the intensity of which will decay exponentially with the speed parameter β. $\boldsymbol{\vartheta}$ is defined as a $(d \times d)$ branching matrix $\boldsymbol{\vartheta} := \vartheta_q^i; q, i \in \{1, .., d\}$ and ϑ_q^i governs the mean increase of the intensity of the process in the dimension i that is raised by events in the dimension q. The impact function $\omega_q^i(\kappa_q)$ measures the increase of intensity in the dimension i that is led by the mark value κ_q in the dimension q. Further details of Marked Hawkes Process and its interpretations can be seen in Liniger (2009).

For the purpose of letting the ground intensity $\lambda_g^i(t)$ to be stationary so that the process $\{N_t^i\}_{i=1,\dots,d}$ has stationary increments, we assume the following conditions

(1 and 2) hold:

Condition 1 *The branching matrix $\boldsymbol{\vartheta}$ satisfies the condition that its spectral radius or $L^1 - norm$ is smaller than 1 ($||\boldsymbol{\vartheta}|| < 1$).*

Condition 2 *The non-negative decay function $\phi(\cdot)$ satisfies $\int_0^\infty \phi^i(t)dt = 1$, and $\int_0^\infty t\phi^i(t)dt < \infty, \forall i \in \{1, ..., d\}$.*

See a more detailed definition and proof for the stability of a marked Hawkes process in Daley et al. (2003) with which most of our mathematical setup is in line.

The marked Hawkes process was proposed and originally applied to the field of seismology and neurophysiology, genome analysis and criminological research (see e.g. Brillinger 1988, Reynaud-Bouret & Schbath 2010, Mohler et al. 2011). It has attracted many studies in the finance field in the recent decade.

Aït-Sahalia et al. (2015) incorporate a multivariate Hawkes process into an Itô's semimartingale and study the cross-impact of international stock markets. Li & Zinna (2018) propose a pricing model for asset returns and variance swap rates, and find evidence of self-exciting behaviours of jumps. Lee & Seo (2017) apply a MMPH to take into account both the impact of marked values and periodicity. They assume a symmetric structure in their study of price and variance. They find that intraday periodicity is an important aspect of the modelling of price and variance dynamics. Fulop et al. (2015) also find evidence of self-exciting jumps, especially during a financial crisis. Maneesoonthorn et al. (2017) use high-frequency data to detect jumps in daily asset prices, and show that self-excitation is apparent in these estimates. Their modelling uses daily-frequency data.

1.3 Motivations and Contributions

This thesis considers a stochastic volatility model in the high-frequency setting. This is motivated by potential intraday dynamics of jump dependency. Specifically, a jump can raise the probability of a jump happening in the following few hours, but

this may not show up in daily movements. A potential explanation for this intraday dependency of jumps comes from information asymmetry (Grossman 1976). Suppose a publicly-listed company releases information which tend to be negative and may deteriorate the company's stock price. The most informed investors would liquidate the company's stock from their portfolio, which may create a negative jump in company's stock price. Less informed investors may follow suite, potentially producing further negative jumps clustering together. Thus, modelling jump clustering in an intraday level can be important.

Based on the potential intraday dynamics of jump dependency, Chapter 2 formalise this idea and propose a stochastic volatility model embedded with a three-dimensional marked Hawkes process. It also considers the impact of intraday periodicity. To our best knowledge, we are the first to propose a model for high-frequency asset prices with jump dependencies. In the empirical studies, we retrieve five-minutes data of five US individual stocks and a broad index (S&P 500). Regarding the model estimation, we combine the non-parametric estimations with the Bayesian estimation to estimate the model. This is to increase the stability of Bayesian Markov chain Monte Carlo algorithm. Our results indicate evident jump clustering in both individual and index data. We also show the importance of jump sizes in modelling and an asymmetric impact structure among three dimensions of jump (positive, negative price jumps and variance jumps). In addition, we show that jump clustering behaviours can be 'hidden' by intraday periodicity.

We evaluate our model proposed in comparison to benchmark models in two ways. First, we report deviance information criteria, and the Bayes Factor. Results are shown to favour the modelling price and variance jumps with the marked Hawkes process. Second, we define a cluster of jumps and evaluate models we consider in terms of their ability to reproduce characteristics found in the data - namely mean and standard deviation of cluster's length. This emphasises the models' capability to capture aspects of particular importance. Consequently, the model with marked Hawkes process generally outperforms the other models in terms of reproducing

these characteristics.

Chapter 3 investigates the contribution of including jump clustering to forecasts on variance and risk measures based on the model in Chapter 2. The model estimation remains unchanged in the in-sample. During the out-of-sample, we apply a particle filter to iteratively estimate variance at each state and forward simulate return and variance distributions, from which we retrieve the forecasts of variance and two risk measures, namely value-at-risk (VaR) and expected shortfall (ES). We also consider a range of forecast horizons from five minutes to one trading day. Furthermore, we apply a Kalman filter to correct forecast errors caused by inconsistently estimated parameters. Through a simulation study, we show that the micro-structure noise in high-frequency data can lead to inconsistent estimation of parameters.

In the empirical work, we examine an individual stock and an index data (Apple Inc. and S&P 500 Index). We first examine the performance of high-frequency variance forecasts in terms of mean square error, forecast bias, R-square from a Mincer-Zarnowitz regression and Diebold-Mariano test of our model against others. We also look at forecast performance after the arrivals of different types of jumps. Forecast results favour the model with jump clustering features in general. We also find the data associated with more jumps reports comparatively bigger improvements in variance forecasting when including jump clustering in the model. In addition, results show that negative jump clustering contributes more in forecasting and performance is relatively better with forecasting horizons of less than two hours.

We set the forecast horizon to one trading day and forecast on daily realised volatility. We benchmark two popular classes of realised volatility forecasting models in current literature, namely heterogeneous autoregressive (HAR) family models and realised GARCH family models. The forecasting results indicate the superiority of models incorporating jump clustering features.

Lastly, we take quantiles of predicted return distributions and retrieve predicted VaR and ES. In the evaluation of out-of-sample forecasting, we adopt a loss function

proposed by Fissler & Ziegel (2016) to jointly assess forecasts of VaR and ES. We find that the inclusion of jump clustering in a model can reduce the expected loss of VaR and ES by up to 15%. Similar to high-frequency variance forecasting, we also examine expected losses after the arrivals of different types of jumps across a range of forecast horizons. The results suggest that the model with jump clustering features outperform others, especially after negative price jumps and when the forecast horizon is less than one hour. In addition, we adopt two backtesting techniques on VaR and ES forecasts, namely the DQ (dynamic quantile) test proposed by Engle & Manganelli (2004) and a DES (dynamic expected shortfall) test proposed by Patton et al. (2019). The backtesting results suggest that the forecasts of VaR and ES by our model are least likely serially correlated.

Chapter 4 use the same model but in a more parsimonious form. The fourth chapter aims to apply the model on Bitcoin data and predict price jumps in the data. Similar to the third chapter, we estimate the static parameters in the in-sample period by Bayesian MCMC and fix them at their posterior mean in the out-of-sample period. In the out-of-sample, we adopt a particle filtering to iteratively estimate variance at each state and conduct one-period-ahead probabilistic forecasting on both positive and negative jumps. We assess the probabilistic forecasting by a continuous ranked probability score (CRPS). We take two measures on true values of jumps, which are Bayesian estimated jumps and non-parametrically estimated jumps. In our empirical studies on Bitcoin data in a range of frequencies, CRPS and associated DM test results show that our forecasting framework outperform two other benchmark models, especially when data frequency is higher than 30 minutes.

We further propose a new statistic that takes the differences between the predicted probability of positive jumps and that of negative jumps. We show that the empirical distribution of this statistic has evident tails and those tails have strong indications on Bitcoin returns. Specifically, returns are more likely associated with positive variations when the predicted probability of positive jumps is higher than that of negative jumps, and vice versa. We further demonstrate this point by a

regression analysis. In addition, we propose a trading strategy to long the returns on the right tail of the statistic and short the returns on the left tail of the statistic. We consider the impact of transaction costs, trading gaps and liquidity costs. The Sharpe ratio of our strategy can reach up to 4.36 with costs. We also find that the exceptional performance of the Sharpe ratio stems from a low standard deviation of strategy returns. The performance of maximum drawdown further shows our jump prediction framework's ability to mitigate downside risks. We further conduct a battery of robustness checks, including higher transaction costs, longer trading gaps, a range of performance evaluation ratios and assessing if the strategy works better during different periods.

This thesis will be organised as follow. Chapter 2 studies stochastic volatility model embedded with a marked Hawkes process and its application in high-frequency data. Chapter 3 conducts high-frequency forecasting on variance and two other risk measures based on the model. Chapter 4 studies forecasting return jumps in Bitcoin data with a more parsimonious model. Chapter 5 summarises the findings of the thesis and discusses opportunities for future research.

To improve the readability of the thesis, we make each chapter self-contained. We introduce models and parameters in each chapter, but we strive to align all notations throughout the thesis whenever possible.

Chapter 2

Modelling Price and Variance Jump Clustering Using the Marked Hawkes Process

2.1 Introduction

Understanding the behaviour of large market movements or jumps in asset pricing and its variance is essential to risk management. There are many studies that examine the importance of including both price and volatility jumps in asset pricing models (see for example, Eraker (2004), Asgharian & Bengtsson (2006), and Barndorff-Nielsen & Shephard (2006)). These studies often assume serial independence of the jump components, and use daily data in their empirical studies. However, less is known about the extent to which one jump presages subsequent jumps, especially at the intraday level. In this chapter, we study the clustering behaviour of jumps using intraday high-frequency data and a marked Hawkes process (MHP) embedded in a bivariate jump-diffusion model. We find evidence of self-excitation behaviour of jumps in both individual stocks and in a broad equity index, and that jumps in prices and volatilities impact one another in an asymmetric manner. We

also find evidence that the magnitudes of the jumps matter. Further, we demonstrate that the inclusion of jump parameters in a model significantly improves the fit of the model. Lastly, we simulate jump intensity under our model settings and study the cluster characteristics where we the find model with a MHP outperform other benchmark models in terms of reproducing some characteristics of original data.

A potential explanation of why jumps might cluster comes from information asymmetry (Grossman 1976). Suppose a publicly-listed company's financial report indicates lower profitability. The most informed trader or investor would liquidate the company's stock from their portfolio, which may have a negative effect on the company's stock price. Less informed investors may follow suite, potentially producing negative variations or negative jumps clustering together. Lee (2012) also find evidence of jumps in stock market that is caused by information releases both in macro level and firm level. For these reasons one might want to allow for jump clustering in traditional asset pricing models, especially in recent times given the higher than normal volatility in financial markets.

Some previous studies provide models of events clustering, where they model events by using a self-exciting process or Hawkes process (HP) (as proposed by Hawkes (1971*a*,*b*)). HP differs from a Poisson process, where events arrive randomly and independently of each other. HP relaxes the assumption of independent arrivals of events, and allows the underlying intensity of events to depend on past events. Additionally, HP can be extended to marked Hawkes process (MHP) and multivariate marked Hawkes process (MMHP) (e.g., Liniger (2009). These models emphasise the impact of marked values (jump sizes) by allowing the intensities of events to depend on occurrences of past events. In addition, intensities are allowed to depend on the marked values attached to *other* past events (e.g., price jumps may depend on variance jumps).

In recent studies there have been many financial applications. Aït-Sahalia et al. (2015) apply a multivariate HP in studying jumps among international markets (see

also Gresnigt et al. (2016) and Lee & Seo (2022)). Lee & Seo (2017) apply a MMHP to take both the impact of marked values and periodicity into account. They assume a symmetric structure in their study of price and variance. They find that intraday periodicity[1] is an important aspect of the modelling of price and variance dynamics (also see studies by Andersen & Bollerslev (1997) and Boudt et al. (2011)). In Lee & Seo (2017)'s study, they cut off the first and last 30 minutes in every trading day to reduce intraday periodicity, and study the interaction between price and variance jumps. These studies provide a number of valuable insights regarding the the application of Hawkes models. However they use non-parametric methods to filter out financial market jumps. We are interested in incorporating the jump clustering feature into a general stochastic volatility model.

A strand of literature considers parametric models of price and variance dynamics for processes comprising both continuous and jump components. They accommodate Poisson or Lévy processes in a continuous time semimartingale, and commonly adopt the assumption that jumps arrive randomly and increments are mutually independent (see e.g., Merton 1976, Duffie et al. 2000, Eraker 2004). More recently, potential interactions between price and variance jumps have been studied by Jacod & Todorov (2010) and Bandi & Reno (2016). Our goal is to investigate whether allowing for jump clustering provides superior models that better capture the characteristics of actual data.

Maneesoonthorn et al. (2017) is similar to our study in some respects. They use high-frequency (intraday) data to detect jumps and calculate variance estimates, and show that self-excitation is apparent in these estimates. However, their modelling uses daily-frequency data. Consequently they may miss intraday clustering: a jump can raise the probability of a jump happening in the following few hours, but this may not show up in daily movements.

[1] The variance of returns can vary over a trading day, and the variation pattern tends to be highly correlated with trading volume which is often higher during market opens and closes (Andersen & Bollerslev 1997).

Our interest is investigating the stochastic volatility model in the high-frequency setting: a jump occurring in one day might in fact comprise a number of jumps within that day. At a daily level, any interdependencies in the intraday jumps would not be evident. Consider again a company disclosing negative news during trading hours. Traders are able to liquidate their assets very quickly, perhaps within the same day, establishing intraday movements in prices and volatilities. Hence jump clustering or their self-exciting behaviours, will be apparent in high-frequency data and may diminish on a daily level. Other aspects may be of interest too: whether jump size is a determinant of clustering behaviour, and the nature of the relationship between the magnitude of the jumps and the effect on the future intensity of jumps.

Our main contribution is to study the intraday dynamics of jump clustering within a general continuous-time asset pricing model. Our methodological contribution it to embed a MMHP in a price and variance state space model, and simultaneously estimate variances, jump magnitudes, the effects price and variance jumps have on each other, etc. In addition, we consider intraday *periodicity*, and the potential this has to 'hide' clustering behaviour. In the estimation, we estimate the model in a hybrid fashion of non-parametric estimation and Bayesian inference to increase stability of Bayesian estimation algorithm.

In our empirical work, we find that both high-frequency stocks and an index exhibit self-exciting features. We also find that allowing for intraday periodicity reduces the number of variance jumps to less than 10% of the number that is found in the model when the periodicity is ignored. Most variance jumps happen during the market open and close and are not identified as jumps when we allow for intraday periodicity. Our model also quantifies jump intensity, for example, using S&P 500 data, we show jump intensity or probability of jumps is lower than 1% during a 'peaceful' period and becomes as high as 50% in during a cluster, although it decays quickly in the next few hours. We also show that price and variance jumps interact in an asymmetric fashion. Specifically, negative price jumps are more likely to be produced by past negative jumps rather than positive jumps, while positive

jumps are as likely to be produced by past jumps of either sign. In addition, in some individual stocks, jump sizes are positively correlated with the following jump intensity: larger jumps tend to escalate higher jump intensity in the future.

To judge the support the data lends the various models in our study, we report the DIC (Deviance Information Criteria), and the Bayes Factor. These are shown to favour the modelling of returns and variance with marked HP models.

We also evaluate the range of models we consider in terms of their ability to generate features found in the data - namely aspects of the clustering of jumps. This goes beyond the general fit of the models to the data, and highlights the ability of the models to capture aspects of particular importance. The MMHP generally outperforms the other models in terms of reproducing these characteristics. There is some evidence that the simpler HP is already sufficient to reproduce some features of the clustering of variance jumps.

The findings in this chapter may be relevant to risk managers, high-frequency traders and other practitioners, who may benefit from knowledge of the inter-dependencies of jumps.

The chapter is organised as follows. Section 2 contains the theoretical framework. Section 3 presents the estimation method, along with jump detection methods. Section 4 discusses the empirical findings. Section 5 we describes the simulation tests. We conclude in section 6. Some technical results are confined to an appendix.

2.2 Theoretical Setup

In this section, we introduce a continuous-time price and variance jump-diffusion process embedded with a MMHP, details of the MMHP, and the discretised form of the model.

2.2.1 Continuous-Time Representation of Price and Variance Process

We let P_t be natural logarithm of asset prices and V_t be the variance at time t, and consider the following jump diffusion process:

$$dP_t \cdot s_t = \mu dt + \sqrt{V_t} dW_t^P + \xi_t^{P+} dN_t^{P+} + \xi_t^{P-} dN_t^{P-} \tag{2.1}$$

$$dV_t = \gamma(\theta - V_t)dt + \sigma_V \sqrt{V_t} dW_t^V + \xi_t^V dN_t^V, \tag{2.2}$$

where s_t is a periodic component, μ is a drift term, γ and θ denotes the mean reversion speed and long-run variance mean, respectively, and σ_V refers to the volatility of volatility. W_t^P and W_t^V are Wiener processes of return and variance respectively, and we let increments of them be correlated $E(dW_t^P, dW_t^V) = \rho dt$. V_t denotes the variance in the process. In terms of the jump components in the processes, we separate price jumps into positive and negative groups, however, we only consider positive variance jumps in our study since decrease of variance is captured by γ and θ. For ease of estimation, we let sizes of price jumps follow a normal distribution $\xi_t^P \sim N(\mu_P, \sigma_P)$ and those of variance jumps follow an exponential distribution with mean μ_V, $\xi_t^V \sim exp(\mu_V)$. In addition, $\{\xi_t^{P+}, \xi_t^{P-}, \xi_t^V\}$ denotes the size of the jumps. In terms of jump components, we employ the following three-dimensional marked Hawkes process to present the jump components in the processes:

$$P(N_t^i = 1) = \lambda_t^i dt, \quad i = \{P+, P-, V\}, \tag{2.3}$$

where λ_t^i denotes the stochastic intensity of the counting process which is defined as a MMHP in Equation (2.4). This constitutes a departure from the literature. The continuous-time asset pricing model typically assumes a constant stochastic intensity (see e.g., Duffie et al. 2000, Eraker 2004, Bandi & Reno 2016). Our model alows mutually dependent jump intensities, which may be correlated with jump size. Another departure from the literature is our use of intraday data to investigate the dynamics of price, variance and jump clustering, as opposed to the use of daily data in most previous studies.

2.2.2 A Three-Dimensional Marked Hawkes Process

Here we introduce a MMHP, the corresponding intensity of the jump processes in Eq. (2.3) is given by:

$$\lambda_t^i = \lambda_0^i + \sum_{q=1}^{d} \vartheta_q^i \int_{[0,t)\times\Xi} \phi_q^i (t-s)\, \omega_q^i(\xi_q) N_q (ds \times d\xi), \tag{2.4}$$

where λ_0^i denotes the immigration[2] intensity of dimension i (three dimensions are indicated in Eq. (2.3)) which is constant; $\boldsymbol{\vartheta}$ is a branching coefficient matrix; ϕ_q^i is a decay function such that $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ and we assume a exponential decaying kernel $\phi_q^i(t-s) = \beta_q^i e^{-\beta_q^i (t-s)}, \beta_q^i > 0$,. $\omega_q^i(\xi_q)$ is an impact function of jump sizes, such that $\omega : \mathbb{R} \to \mathbb{R}_+$.

Remark 1 (Interpretations) *Under a general Hawkes process, the immigrants arrive as a Poisson process with immigration intensity λ_0^i and jump sizes ξ. They further produce the arrivals of further jumps, the intensity of which will decay exponentially with the speed parameter β. $\boldsymbol{\vartheta}$ is defined as a (3×3) branching matrix $\boldsymbol{\vartheta} := \vartheta_q^i; q, i \in \{P+, P-, V\}$ and ϑ_q^i governs the mean increase of intensity to the process in dimension i that is produced by jumps in dimension q. The impact function $\omega_q^i(\xi_q)$ measures the increase of intensity in dimension i that is led by the jump sizes ξ_q in the dimension q. Further details of MMHP and its interpretations can be seen in Liniger (2009).*

Remark 2 (A Simple Example) *For a simple example of MMHP, given an occurrence of jump in dimension q (size: ξ_q), the underlying intensity of dimension i will increase from λ_0^i to $\lambda_0^i + \vartheta_q^i \cdot \omega_q^i(\xi_q)$. Then, the incremental part of the intensity $\vartheta_q^i \cdot \omega_q^i(\xi_q)$ will decay at a speed of $\phi_q^i(dt) = \beta_q^i e^{-\beta_q^i \cdot dt}$ for every dt.*

[2]We denote jumps that arrive when $\lambda_t^i = \lambda_0^i$ as immigrant jumps or immigrants. They are also first jumps in clusters of jumps. The other subsequent jumps arriving when $\lambda_t^i > \lambda_0^i$ in clusters can be seen as being produced by the immigrant jump. We use this notation throughout the thesis.

See a more detailed definition and stationary assumptions of a MMHP in Daley et al. (2003) with which most of our mathematical setup is in line.

Remark 3 (Impact Functions) *Inspired by Liniger (2009), we normalised our impact functions to satisfy some stationary conditions, denoting $\tilde{\omega}_q(\cdot)$ as the impact function before being normalised. The normalised impact function is given by:*

$$\omega_q^i(\xi_q) = \frac{\tilde{\omega}_q^i(\xi_q)}{\mathbb{E}[\tilde{\omega}_q^i(\boldsymbol{\xi})]} \tag{2.5}$$

and we consider four impact functions (before normalisation) as. follow:

$$\begin{aligned} \tilde{\omega}_q^i(\xi_q) &= \tilde{\alpha}_q^i + \tilde{\beta}_q^i \xi_q + \tilde{\gamma}_q^i \xi_q^2 \quad (I) \\ \tilde{\omega}_q^i(\xi_q) &= \xi_q^{\tilde{\alpha}_q^i} \quad (II) \\ \tilde{\omega}_q^i(\xi_q) &= e^{\tilde{\alpha}_q^i \xi_q} \quad (III) \\ \tilde{\omega}_q^i(\xi_q) &= \tilde{\alpha}_q^i + \tilde{\beta}_q^i log(1+\xi_q) \quad (IV) \end{aligned} \tag{2.6}$$

We assume the parameters in the above impact functions $\{\alpha, \beta, \gamma\}$ satisfy those within each function, as at least one of the parameters is strictly positive. Also noticeably, the jump sizes have no impact on intensity processes when $\omega(\xi) \equiv 1$. In addition, we assume an identical independent distribution of jump sizes.

2.2.3 Discretised Form of Return and Variance Processes

Similar to previous studies, we apply an Euler discretisation to the processes with $\Delta t = \frac{1}{78 \times 252}$ (equivalent to five minutes) and obtain the following forms:

$$(P_t - P_{t-1}) s_t = \mu + \sqrt{V_{t-1}} \epsilon_t^P + \xi_t^{P+} \Delta J_t^{P+} + \xi_t^{P-} \Delta J_t^{P-} \tag{2.7}$$

$$V_t = \alpha_v + (1 + \beta_v) V_{t-1} + \sigma_V \sqrt{V_{t-1}} \epsilon_t^V + \xi_t^V \Delta J_t^V, \tag{2.8}$$

where P_t and V_t denotes the logarithm of the asset price and the variance at time t, s_t is a periodic component and $s_t = I_{t,\tau} \cdot f_\tau$, where $\tau = 1, 2, ..., 78$ and $I_{t,\tau}$ is a time

indicator when τ correspond to time t, $I_{t,\tau} = 1$.[3] μ denotes the drift term, and σ_V denotes the volatility of volatility. The mean reversion speed and level of variance are translated to α_v and β_v, where $\alpha_v = \gamma\theta$ and $\beta_v = -\gamma$ (γ, θ are mean reversion speed and level of variance, see Section 2.3), ϵ_t^P and ϵ_t^V are two random variables that follow a normal distribution $N(0,1)$ with correlation $corr(\epsilon_t^P, \epsilon_t^V) = \rho$. $\xi_t^i, i = \{P+, P-, V\}$ denotes jump magnitudes, and we specify them to follow distributions as follows:

$$\xi_t^{P+} \sim N(\mu_{P+}, \sigma_{P+})\mathbf{1}_{\xi^{P+}>0} \tag{2.9}$$

$$\xi_t^{P-} \sim N(\mu_{P-}, \sigma_{P-})\mathbf{1}_{\xi^{P-}<0} \tag{2.10}$$

$$\xi_t^V \sim exp(\mu_V), \tag{2.11}$$

where the sizes of price jumps follow truncated normal distributions and variance jump sizes follow a exponential distribution with a parameter μ_V, while $\Delta J_t^i = J_t^i - J_{t-1}^i, i = \{P+, P-, V\}$ is a Bernoulli random variable with the corresponding time-varying intensity λ_t^i, so:

$$\Delta J_t^i \sim Bernoulli(\lambda_t^i), \quad i = \{P+, P-, V\}, \tag{2.12}$$

$\Delta J_t^i = 1$ can be viewed as an occurrence of a jump and λ_t^i can be regarded as the probability of a jump happening at time t. Here is where we make distinctions between the models, in the simplest and the most adopted setting where jumps are assumed to follow a Poisson process, underlying intensity λ_t^i is constant at $\lambda_t^i \equiv \lambda_0^i$, such that the probability of a jump happening is the same over time and jumps arrive independently. For a multivariate HP, without considering the impact of marks, the intensity process is specified as follows:

$$\lambda^i(t) = \lambda_0^i + \sum_q \vartheta_q^i \sum_{0<s<t} \phi^i(t-s), \quad i, q = \{P+, P-, V\}, \tag{2.13}$$

[3]We employ 5-minutes data, and therefore, have 78 observations in a trading day, our time t is indexing every 5 minutes, and when τ correspond to time t, we adjust returns by the periodic component f_τ.

where ϑ_q^i is a branching matrix and $\phi^i(\cdot)$ is an exponential decaying kernel (see details of interpretations in Remark 1 and 2). In addition, by adding an impact function on jump sizes $\omega_q^i(\dot{)}$, a MMHP can be introduced to define the intensity of dimension i at time t as follows:

$$\lambda^i(t) = \lambda_0^i + \sum_q \vartheta_q^i \sum_{0<s<t} \phi^i(t-s)\omega_q^i(\xi_q), \quad i, q = \{P+, P-, V\}, \tag{2.14}$$

where $\omega_q^i(\cdot)$ is the impact function and $\omega_q^i(\cdot) \equiv 1$ in the previous multivariate HP setting without the impact of jump size. We adopt four different impact functions (see Remark 3). By adding components of the right-hand side of Equation (2.14) in addition to λ_0^i, intensity or probability of a jump happening is allowed to depend on past jumps through the branching matrix ϑ_q^i and jump sizes through impact functions $\omega_q^i(\cdot)$.

Unlike previous studies (e.g., SVCJ model proposed by Duffie et al. (2000)), we do not specify a dependent structure between the size of the price and variance jump; instead, they are assumed to be identically independently distributed and treated as latent variables in the estimation. We focus on how the size of jumps estimated impact underlying intensity of jumps.

2.3 Parameter Estimation

In this section, we discuss our estimation approaches. We estimate our model in a hybrid fashion of non-parametric estimation and Bayesian inference (Markov Chain Monte Carlo (MCMC)). Specifically, we estimated jumps $\{\Delta J_t^{P+}, \Delta J_t^{P-}, \Delta J_t^V\}$ and the periodic component s_t non-parametrically, and apply them in Bayesian inference. Reason for this is to increase stability of MCMC algorithm. Due to the inclusion of MMHP, we have more static parameters to be estimated jointly than traditional stochastic volatility models, for example, we have 78 parameters in s_t and additional parameters in the branching matrix and impact function, plus they are all in 3×3 dimensions. We found parameters becomes very unstable and converge very slowly

if treating jumps as latent variables, but this can be improved when estimate jumps and the periodic component non-parametrically. The question left is the potential measurement errors of this approach. We conduct a simple simulation study on this issue and results shows our estimation strategies is reasonably powerful[4].

2.3.1 Price Jump Detection

We denote $r_t = (P_t - P_{t-1}) \cdot s_t$ as returns adjusted for intraday periodicity. Our non-parametric jump filtering method is mainly based on Mancini et al. (2015) and Figueroa-López & Mancini (2019). We identify a jump at time t, $J_t = 1$, when the squared return is greater than a threshold, $r_t^2 > \widehat{V_t^2} \cdot 2\Delta log\frac{1}{\Delta}$. $\widehat{V_t^2}$ is a non-parametric estimator of spot variance based on pre-truncated returns:

$$\widehat{V_t^2} = \sum_{\tilde{t}=1}^{n} f_h(t-\tilde{t}) r_t^2 \cdot \mathbb{1}_{\{r_t^2 \leq 9\Delta^{0.99}\}} \tag{2.15}$$

where $f_h(\cdot)$ is weight function, $f_h(t) = \frac{1}{h} \cdot \frac{e^{-|t/h|}}{2}$ with a bandwidth $h = 200\Delta$ for simplicity. The idea of this filtering is to extract those standardised squared returns $(r_t\Delta/\widehat{V_t^2})$ which is not generated by a Brownian motion, whose absolute value is greater than the threshold $\sqrt{2log(1/\Delta)}$.

2.3.2 Variance Jumps Detection

In testing variance jumps, we set our null hypothesis $H_0 : |\widehat{V_t^2} - \widehat{V_{t-}^2}| = 0$ against the alternative hypothesis that there is a variance jump $|\widehat{V_t^2} - \widehat{V_{t-}^2}| > 0$. So, the test statistic in this test should be a function on the difference of spot variance

[4]We simulate data using our model with impact function IV, we set parameters at posterior mean of estimated parameters using S&P 500 Index data, and simulate 10000 times. We find our price jump detection has a power at approximately 71% in average, and variance jump at 66% (percentages of simulated jumps being correctly detected). We also find the estimated periodic component s_t is approximately overlap with true value with an error of less than $\pm 7.6\%$. We choose not to present other simulation results (e.g. model fitness of simulated data) since it may strongly rely on data generating process, but examining power of estimation approach is useful for us.

$f(\widehat{V_t^2}, \widehat{V_{t-}^2})$. Following Jacod & Todorov (2010), we construct our test statistics as follows:

$$\mathcal{L}^v(t) = 2log(\frac{1}{2}(\widehat{V_t^2} + \widehat{V_{t-}^2})) - log(\widehat{V_t^2}) - log(\widehat{V_{t-}^2}) \tag{2.16}$$

and $n^b\mathcal{L}^v(t) \to \mathcal{X}_1^2$, where n denotes the number of observations and $b = \frac{1}{2} - \delta$, where $\delta = \frac{1}{78\times252}$. The temporal variance estimator $\widehat{V_t^2}$ is specified in Eq. 2.15.

2.3.3 Intraday periodic Effects

High-frequency financial data displays intraday periodicity, (see for example, Hecq et al. (2012); Andersen et al. (2019)). As is shown in Figure 2.1, the mean absolute return of the S&P 500 shows a clear U-shape, the mean absolute value of overnight returns (first five-minute return of a day) is much higher than returns at other time. We also find that most of returns during the market open are identified as jumps, as shown in Figure 2.2 (up left). The duration between two jumps mostly appears to be 78 units of times (δ) or its multiples, which could imply most price jumps occur at the market openings.

[INSERT FIGURE 2.1 ABOUT HERE]

[INSERT FIGURE 2.2 ABOUT HERE]

Boudt et al. (2011) show that taking intraday periodicity into account can improve the overall accuracy of jump detection. Therefore, we adopt a weighted standard deviation (WSD) estimator proposed by Boudt et al. (2011), which is based on a shortest half scale estimator proposed by Rousseeuw & Leroy (1988). Using this approach we define the order statistics of returns $\bar{r}_{(1),i} \leq \bar{r}_{(2),i} \leq ... \leq \bar{r}_{(T_i),i}$. The shortest half-scale statistics are determined as follows:

$$ShortH_i = 0.741 \cdot min\{\bar{r}_{(h_i),i} - \bar{r}_{(1),i}, ..., \bar{r}_{(T_i),i} - \bar{r}_{(T_i-h_i+1),i}\}, \tag{2.17}$$

where $h_i = \frac{T_i}{2} + 1$, which makes the statistics essentially minimum differences among all of the return's halves. The shortest half-scale estimator for periodicity is given

by:

$$\hat{f}_i^{ShortH} = \frac{ShortH_i}{\frac{1}{T}\sum_{j=1}^{T} ShortH_j^2}, \tag{2.18}$$

where T denotes the number of observations within a day and the WSD estimator can be obtained as follow:

$$\hat{f}_i^{WSD} = \frac{WSD_i}{\frac{1}{T}\sum_{j=1}^{T} WSD_j^2}, \tag{2.19}$$

where:

$$WSD_j = \sqrt{1.081 \cdot \frac{\sum_{k=1}^{T_j} w_{k,j}\bar{r}_{k,j}^2}{\sum_{k=1}^{T_j} w_{k,j}}}, \tag{2.20}$$

$w_{k,j} = I(\bar{r}_{k,j}/\hat{f}_i^{ShortH})$ is a weight function with a identification function $I(\cdot)$, such that $I(x) = 1$ if $x \leq 6.635$ and 0 otherwise. We use the WSD estimator ($\hat{f}_i^{WSD}$) as a periodic component (s_t) estimator.

Looking back to interarrival times (durations between jumps) after adjusted by intraday periodic effects (de-periodisation) in Figure 2.2 (two figures on the right-hand side), as a comparison, we filter price and variance jumps without considering intraday periodicity [5] (two figures on the left-hand side). After de-periodisation, duration between two jumps generally decreases exponentially over time in the histograms and tends to exhibit a decline in periodicity, while before de-periodisation, interarrival times are mostly lie in 78 or 78's multiples (78 five-minutes interval stands for 6.5 trading hours which is a typical trading day), which imply most jumps occur at market openings and closings.

2.3.4 Bayesian Inference on Parameters

We next conduct a Bayesian inference on price, variance, and jump processes. We perform non-parametric jump tests and estimate the periodic components as known inputs in the inference, and treat jump sizes, the branching matrix and variance

[5] In this chapter, we refer "without considering periodic effects" and "before de-periodisation" to setting the periodic component $s_t \equiv \mathbf{1}$.

$\{\xi^{P+}, \xi^{P-}, \xi^{V}, \boldsymbol{\vartheta}, V_t\}$ as latent variables. We follow Rasmussen (2013) to conduct the estimation of parameters in the Hawkes process. We embed the process in the price and variance dynamics. We denote the static parameter vector as $\Theta = \{\mu, \alpha_v, \beta_v, \sigma_V, \rho, \lambda_0^i, \beta^i, \Phi^i\}$, where Φ^i denotes parameters in the impact function (2.6) and depends on which function is used. We also denote the data we can observe and other non-parametric estimations we derive from the available data as $Y_t = \{P_t, J_t^{P+}, J_t^{P-}, J_t^{V}\}$. Then, joint posterior distribution can be given using a Bayes formula as:

$$P(\Theta, \xi^{P+}, \xi^{P-}, \xi^{V}, \boldsymbol{\vartheta}, V_t | Y_t) \propto P(Y_t | \Theta, \xi^{P+}, \xi^{P-}, \xi^{V}, \boldsymbol{\vartheta}, V_t) P(\xi^{P+}, \xi^{P-}, \xi^{V}, \boldsymbol{\vartheta}, V_t | \Theta) P(\Theta), \tag{2.21}$$

The posterior given in Equation 2.21 is apparently not available in closed form due to the complexity of the processes. Therefore, we adopt a Markov Chain Monte Carlo (MCMC) method to generate a sequence of draws on parameters and latent variables and simulate the posterior. The specification of priors and details of the algorithm are in Appendix A.1.

2.4 Empirical Application

In this section, we introduce our dataset and models. We will firstly report parameters estimated in the price and variance process and MMHP kernels. We then compare the models in terms of their fit to the data, and finally present a test of the ability of the models to reproduce jump clustering characteristics of the actual data.

2.4.1 Data

We retrieved five-minute price data from Bloomberg for four individual stocks across different industries; namely, Apple (AAPL), Boeing (BA), J.P. Morgan (JPM), Coca Cola (KO) and S&P 500 Index from 3/1/2012 to 31/12/2019. We chose these stocks

as they are large firms that have high liquidity, are from different industries, and are likely to be followed by analysts and therefore react to news quickly. This actually put us in a disadvantageous place in terms of showing the usefulness of considering jump clustering, since small-cap and less-liquid stocks potentially exhibit more evident jump clustering behaviours. However, we show that jump clustering also exist in large-cap liquid stocks. We also examine three ETF data, namely iShares FTSE China Index Fund (FCHI), iShares MSCI Spain ETF (IBEX) and SPDR S&P 500 Trust ETF (SPY), full results are in the Appendix. We cleaned the data following the approach in Barndorff-Nielsen et al. (2008). We set the time unit $\delta = \frac{1}{M \cdot 252}$ years where M is number of daily observations. Here we plot the log return, estimated variance, estimated price and variance jump sizes of Apple (AAPL) stock and the S&P 500 Index, respectively, from 2012 to 2019.

[INSERT FIGURE 2.3 ABOUT HERE]

[INSERT FIGURE 2.4 ABOUT HERE]

We filtered price and variance jumps before and after we de-periodise stock returns. Table 2.1 shows that the impact that including intraday periodicity has on the number of jumps detected is mixed. For some individual stocks, the number of price jumps triples while others do not increase at all, and the S&P 500 Index has less price jumps considering intraday periodicity. Also, the reduced number of variance jumps suggest that most variance jumps happen during market opens.

[INSERT TABLE 2.1 ABOUT HERE]

2.4.2 Parameters Estimated in Price and Variance Process

We estimate parameters in (2.7) and (2.8), but as a comparison, we model the intensity of jumps $\{J_t^{P+}, J_t^{P-}, J_t^V\}$ differently. We specify seven models ($\mathcal{M}_1$ to $\mathcal{M}_7$) in Table 2.2 to present different methods in modelling jump intensity. $\mathcal{M}_1$ is

our benchmark model, SVCJ model proposed by Duffie et al. (2000) [6]. $\mathcal{M}_2$ and $\mathcal{M}_3$ are models with jump clustering effect modelled by multivariate HP, but $\mathcal{M}_3$ consider intraday periodicity. $\mathcal{M}_4$ to $\mathcal{M}_7$ model jumps by MMHP with different impact functions.

[INSERT TABLE 2.2 ABOUT HERE]

In Table 2.3, we report parameters we estimate in the price and variance processes with $\mathcal{M}_1$. We would skip these parameters in other models since we find little differences, our interests are more in parameters of jump components:

[INSERT TABLE 2.3 ABOUT HERE]

There are some discrepancies in Table 2.3 between our results and the results in other similar models examined by, for example, Asgharian & Bengtsson (2006), who used daily data. This is probably due to the high-frequency data we use which is more turbulent and contaminated by microstructure noise, and the drift term and mean reversion level can be masked by noise and become less significant. Note also that our correlation ρ is also lower than that estimated using daily data.

2.4.3 Significance of jump clustering with and without Intraday periodicity

In this subsection, we report parameters of Hawkes kernels we estimate in $\mathcal{M}_2$ and $\mathcal{M}_3$. We set the impact function to $\omega_q(\xi_q) \equiv 1$, where we assume that jump sizes do not have impact on future jump intensity. A summary of the results is in Table 2.4 with full results reported in the Appendix. In Table 2.4, we give the posterior mean of the parameters, and number of which are significant. For price jumps, there is a clear trend showing that without periodicity ($\mathcal{M}_2$) as indicated by

[6] They assume constant jump intensity but let jump sizes to be correlated: $\xi_t^P \sim \mathcal{N}(\mu_J + \rho_J \xi_t^V, \sigma_J)$

$\{\vartheta^{p+}_{p+}, \vartheta^{p+}_{p-}, \vartheta^{p-}_{p+}, \vartheta^{p-}_{p-}\}$, are significantly greater than 0, but this self-excitation feature only appears in individual stocks data, not the S&P 500 Index. This result provides supports for previous studies such as Foschi et al. (2019). However, with periodic components ($\mathcal{M}_3$), price jumps in the S&P 500 also have the self-exciting feature. In addition, we find that jumps in variance, to some extent, can produce future price jumps after considering periodicity. Similarly, variance jumps also become self-excited and can be produced by price jumps after considering periodicity. We also find that the decaying speed of some variance jumps is faster than others when we examine the values of $\{\beta^v_v, \beta^v_{p-}, \beta^v_{p+}\}$.

The results can also indicate the asymmetry of the branching coefficient matrix $\boldsymbol{\vartheta}$. Specifically, posterior mean of ϑ^{p+}_{p-} is not statistically significantly different from that of ϑ^{p+}_{p+} (0.029 and 0.04 in $\mathcal{M}_2$; 0.159 and 0.172 in $\mathcal{M}_3$). This shows that there is no clear evidence whether positive jumps is produced by positive or negative price jumps. However, posterior mean of ϑ^{p-}_{p-} is statistically significantly greater than ϑ^{p-}_{p+} (0.042 and 0.019 in model $\mathcal{M}_2$; 0.216 and 0.108 in $\mathcal{M}_3$). This suggest negative jumps are more likely produced by itself rather than produced by positive jumps.

Additionally, the impact of variance jump on other types of jump is mixed. Although it seems the intensity contributed by variance jumps is higher, it also is often associated with faster decaying speed. This result is different from that proposed by Maneesoonthorn et al. (2017), who find that variance jumps persist longer and price jumps are short-lived. However, they use daily data while we use five-minute data. We also suspect that another reason for this contradictory findings is very small number of variance jumps filtered out compared to the number of price jumps after considering periodicity. We find these small number of variance jumps are just scattered around price jumps by chance, which may also be the reason for the fast decaying speed because it is not really signalling future jumps. Additionally, variance jumps are less likely self-excited but more likely produced by price jumps.

2.4.4 Significance of Impact from Jump Sizes

We next examine the impacts of jump sizes. The results of the jump-size-related parameters, $\tilde{\beta}$ and $\tilde{\gamma}$ in $\mathcal{M}_4$ and $\tilde{\beta}$ in $\mathcal{M}_7$ are also reported in the table 2.4. These parameters are from impact functions (see Equation 2.6) of MMHPs.

In table 2.4, some of the coefficients of jump size are significantly great than 0, and they are associated with high decaying speed (β). For example, given $\beta = 0.5$, the intensity heightened by past jumps will decay to 5% of the original after five time units (30 minutes), and it will decay to 0.3% within the same time if $\beta = 1$. However, the percentage will be 41% if $\beta = 0.15$. Therefore, the value of β should also be taken into account when judging whether a parameter is statistically significant or meaningful in generating future jumps.

As can be seen in Table 2.4, in terms of the coefficients of jump sizes, some $\{\tilde{\beta}^{p+}_{p+}, \tilde{\beta}^{p+}_{p-}, \tilde{\beta}^{p-}_{p-}\}$, $\{\tilde{\gamma}^{v}_{p-}, \tilde{\gamma}^{p-}_{v}\}$ in $\mathcal{M}_4$ and $\{\tilde{\beta}^{p+}_{p+}, \tilde{\beta}^{p+}_{p-}, \tilde{\beta}^{p-}_{p-}\}$ in $\mathcal{M}_7$ are significantly positive overall. This may suggest that the underlying intensity of positive jumps can be produced by the occurrence of both positive and negative jumps, and they are positively correlated with sizes in the sense that the larger sizes of price jumps can generate higher intensity in future positive price jumps. With regard to negative jumps, they can also be produced by past positive and negative jumps, but the results here imply that only negative jumps are correlated with jump size in escalating the intensity of negative jumps. Moreover, the interaction between negative price jumps and variance jumps is captured by the coefficients of squared jump sizes $\{\tilde{\gamma}^{v}_{p-}, \tilde{\gamma}^{p-}_{v}\}$, but interestingly, they are not significant in the log transform of jump sizes.

Furthermore, $\mathcal{M}_5$ and $\mathcal{M}_6$ may not illustrate whether jump sizes impact the intensity of jumps in any dimensions. Although the parameters in $\mathcal{M}_5$ are mostly significant, they do not provide any information as to whether there is a 'size effect' or they are self-excited without 'size effect', because it will not become self-excited if the parameters are not significant, which is the case in $\mathcal{M}_6$.

In addition, we use parameters estimated in $\mathcal{M}_4$ to simulate jump intensities and plot them in Figure 2.5, the jump intensity can goes from less than 1% to over 50%. As an example, we plot negative return jump intensities of S&P 500 Index 15 trading days around 3/21/2017 at the bottom, the S&P 500 Index was turbulent during the day and dropped by 1.2% at close. As can be seen from the figure, a cluster of 6 jumps occurred and escalate jump intensity to around 0.2 during the trading day.

[INSERT FIGURE 2.5 ABOUT HERE]

2.4.5 Model Fit

To assess the goodness-of-fit of our models ($\mathcal{M}_1$ to $\mathcal{M}_7$), we apply the deviance information criterion (DIC) of Spiegelhalter et al. (2002) and the Bayes factor of Kass & Raftery (1995). The DIC is calculated by model log-likelihood penalised by model complexity and can be easily applied in a MCMC algorithm; a lower value indicates a better fit of the model. Table 2.5 shows that in the $\mathcal{M}_1$ model, jumps modelled using Poisson process provides the highest DIC, while that using MMHP with impact function (2.6, VII) provides the lowest DIC ($\mathcal{M}_7$), hence the best fit. In addition, three of the four MMHP models ($\mathcal{M}_4$, $\mathcal{M}_5$, $\mathcal{M}_7$) outperform the multivariate-HP model without impact of jump size ($\mathcal{M}_3$). Further, not surprisingly, $\mathcal{M}_6$ with an exponential impact function underperform the $\mathcal{M}_3$ since the parameters in $\mathcal{M}_6$ suggests there are few self-exciting features in the data.

Table 2.6 reports the log values of the Bayes factor across seven models. We find that $\mathcal{M}_1$ underperforms all other models. In addition, the results show a large discrepancy between other models against $\mathcal{M}_2$ and $\mathcal{M}_3$, which suggests the importance of considering intraday periodicity. Additionally, the log Bayes factor of $\mathcal{M}_6$ against $\mathcal{M}_3$ is negative, which is in line with the results for the DIC. However, we cannot tell from the results whether $\mathcal{M}_4$ or $\mathcal{M}_7$ provides better fit. Overall, $\mathcal{M}_7$ presents better results but $\mathcal{M}_4$ prevails for the S&P 500 Index data. The marginal

likelihood computation of data given different models can be seen in Appendix 6.4.

[INSERT TABLE 2.5 ABOUT HERE]

[INSERT TABLE 2.6 ABOUT HERE]

2.5 Clusters and Simulation Test

We next evaluate the models by focusing on their ability to capture features of the data of particular interest. We do this by simulating data from the estimated models and seeing whether these data have the features that characterise the actual data. This approach has been used by Hess & Iwata (1997) and Clements & Krolzig (2004) to assess whether a number of time-series models can reproduce business cycle features. Our approach is in line with that used in Clements & Krolzig (2004).

2.5.1 Cluster

In this section we explain how we define a cluster of jumps. The way we define a cluster of jumps is similar to Foschi et al. (2019). The intuition is that jumps may escalate their underlying intensity to be above a "normal" level for a short period, before the intensity returns to the "normal" level. For example, suppose k jumps occurred, and these k jumps are a cluster. If $k = 1$, there are no jumps in the cluster other than the original one. In this case, the intensity returns to normal before any other jumps occur. Jumps occurring subsequent to this are assumed to belong to a new cluster. The formal definition is as follows.

Definition 1 (Cluster) *Given occurrence of k jumps at time $t_1, ..., t_k$ during the period $[t_1, T]$ starting with a immigrant jump at t_1, these k jumps form a cluster if the intensity: 1) before the occurrence of t_1 jump, is within a range around ground intensity $\lambda_{t_1-} \in [\lambda_0, \tilde{\lambda}]$ and 2) above the threshold $\tilde{\lambda}$ during the period $[t_1, T]$. The $\tilde{\lambda}$ is a tolerance level.*

Remark 4 *Alternatively, we could say the cluster does not exhaust until k^{th} jump $\lambda(t_i) > \tilde{\lambda}$, $\forall t_i \in [t_1, T]$ and the cluster will come to end if the intensity returns to the tolerance level.*

2.5.2 Non-parametric Intensity Estimation

The models provide estimates of parameters from which the underlying intensity of jumps can be calculated, allowing us to determine the average number of jumps in a cluster. However, this would make the determination of the degree of clustering dependent on the model, and would favour those models which by design incorporate this feature over others which do not. To enable a fair comparison between the models, we estimate the feature of interest on the models' simulated output, not as a function of the models' parameters. We use a kernel density estimation (KDE) to estimate underlying the intensity of jumps and then determine the clusters. The KDE method used here is very similar to that used in Section 2.3.2. So, the underlying intensity in dimension i at time t is defined as:

$$\hat{\lambda}_t^i = \sum_{\tau=t-h}^{t-1} K_h(\tau - t)\Delta N_\tau^i, \tag{2.22}$$

where $K_h(\cdot)$ is a kernel function with bandwidth h, such that $\int K_h(x)dx = 1$ and $K_h(x) = \frac{1}{h}K(\frac{x}{h})$. Here, we also take exponential kernel $K(x) = \frac{1}{2}exp(-x)$ and for simplicity, we set $h = 40\Delta t$. In terms of the tolerance level, we take $\tilde{\lambda} = \frac{0.05}{2h}e^{-\frac{1}{h}}$.

Remark 5 *We experiment with different bandwidths. The choice of bandwidth does not unduly affect our results for a reasonable range of values, but clearly too large or too small values do give unreasonable results. For example, given an $h = 200\Delta t$, when jumps arrive at a Poisson rate $\hat{\lambda}_t^i \equiv 0.01$, we find they are clustered with significantly more than one jump per cluster, even though they arrive independently*

by virtue of the Poisson process assumption. Similarly, given an $h = 5\Delta t$, there are no signs of clustering in the outputs of any of the models. Additionally, $\tilde{\lambda} = \frac{0.05}{2h}e^{-\frac{1}{h}}$ is calculated as 5% of the underlying intensity when in the previous time period an immigrant jump occurred.

2.5.3 Characteristics of Interests

We choose the average number of jumps in a cluster, and the standard deviation of the number of jumps in a cluster, as the two characteristics of interest. Clustering is a feature of the data, so we compare the models in this dimension - their ability to match this feature of the data - irrespective of their overall fit. Poisson processes treat occurrences of jumps as independent events, so there should, on average. be only one jump in each cluster, and we would expect such processes to fail to match the data in this regard. On the other hand, multivariate HP and MMHP emphasise clustering of jumps, and are expected to produce clusters with more than one jump. However, unlike HP, MMHP allows jumps with larger sizes to raise higher intensities, which also take a longer time to die away. Hence the distribution of the number of jumps under HP and MMHP can be different, and may allow some discrimination between these two models.

We consider $\mathcal{M}_1$, $\mathcal{M}_3$, $\mathcal{M}_4$ and $\mathcal{M}_7$, and treat these models as the DGP, setting their parameters to their posterior means of the Bayesian model estimation. The feature value from the actual data can then be compared to the empirical distributions of the features from the models. If the feature value on the actual data is 'extreme' relative to a model's distribution, we infer that the model is unable to capture that feature.

Specifically, denote by C_i the i^{th} characteristic (either the first or second moments of the number of jumps). Assume there are J simulations. We rank the characteristic on each replication: $\{C_i^{(1)}...C_i^{(J)}\}$. A confidence interval with a significance level SL is given by $[F_{C_i}^{-1}(\frac{SL}{2}), F_{C_i}^{-1}(1-\frac{SL}{2})]$, where F_{C_i} denotes the empirical cumulative distribution of characteristic C_i. For example, taking $J = 10000$, and

significance level at 5%, the upper and lower limit of the confidence interval will then be $F_{C_i}^{-1}(2.5\%)$, and $F_{C_i}^{-1}(97.5\%)$, confidence interval, thus, is taken as $[C_i^{(250)}, C_i^{(9750)}]$.

2.5.4 Testing Results

Table 2.8 presents the summary statistics for number of jumps in clusters. On average, there are two-to-three jumps per cluster, with more jumps in clusters for some individual stock data. Additionally, there is a clear pattern that the average number of variance jumps is lower than that of price jumps. Also, around a half of the clusters consist of more than one jump. We also collect time spans that clusters of jumps cover and denote it as cluster length, as can be seen in the summary statistic table, cluster length of negative return jumps averagely ranges from 30 to 70 time intervals, in the sense, 2.5 to 6 hours. This imply that a negative return jump can have impact on jump probabilities in the next 2.5 to 6 hours. Additionally, the impact can last up to 408 time intervals standing for around 5 trading days. In addition, it is clear that negative return jumps last longer than positive return jumps and variance jumps. This result implies market practitioners to be careful about market risks after spotting a jump, since jumps can escalate the probability of jumps in the next 2.5 to 6 hours.

We simulate 50000 series of data ($J = 50000$) and each contains 100000 data points, Table 2.9 reports the results from each of the models ($\mathcal{M}_1$, $\mathcal{M}_3$, $\mathcal{M}_4$, $\mathcal{M}_7$). In terms of price jumps, the HP model ($\mathcal{M}_3$) is clearly unable to reproduce the features, while the MMHP-class models ($\mathcal{M}_4$ and $\mathcal{M}_7$) perform better. However, in S&P 500 (SPX) data, $\mathcal{M}_7$ outperforms $\mathcal{M}_4$.

In the case of variance jumps, the results are different. For example, for the JPM and KO data, the HP model ($\mathcal{M}_3$) is able to adequately capture the clustering feature. This is likely due to the smaller average number of jumps, and standard deviation, for variances compared to prices. These results are in line with the empirical results in Section 4 (see Table 2.4) where variance jumps are less likely impacted by their mark values or jump sizes. It is noticeable that, although $\mathcal{M}_4$

and $\mathcal{M}_7$ are both MMHP-type models, $\mathcal{M}_4$ always provides higher estimates of the two features.

[INSERT TABLE 2.8 ABOUT HERE]

[INSERT TABLE 2.9 ABOUT HERE]

2.6 Conclusion

In this chapter, we propose a dynamic bivariate jump-diffusion process, in which jump intensities are modelled by a three-dimensional Marked Hawkes Process to allow the occurrences and sizes of jumps to affect future intensities and thus, capture the clustering features. Unlike other stochastic volatility state-space models that apply daily data, we use intraday high-frequency data. In addition to conducting non-parametric methods on jump detection, we further employ a intraday periodic component in the process. Also, a Bayesian MCMC algorithm is constructed to jointly estimate parameters and latent variables in the model.

We find evidence of strong intraday jump clustering in our empirical study. We find that that the self-excitation feature tends to be hidden by the periodicity. We quantify the changing patterns of jump intensity, and show that jump intensity can rise to over 0.5 after the occurrence of large jumps. We investigate the interactions between positive price jumps, negative price jumps, and variance jumps. These are apparent in the branching coefficient matrix $\boldsymbol{\vartheta}$, which represents the extent to which jumps in one dimension affect intensity in other dimensions. This turns out to be asymmetric. Further, our results from the Marked Hawkes Process suggest that the extent to which jumps inflate future intensities is positively correlated with jump sizes: large jumps tend to escalate the probability of jumps in the near future. We assess the fit of the various models via DIC and the Bayes factor, and find that modelling jumps by the Marked Hawkes Process is preferred by these criteria.

Additionally, we study the numbers of jumps in clusters, in price and variance, and we find a cluster of jumps can cover 2.5 to 6 hours on average. Using the

mean and variance of the number of jumps in clusters, we consider the ability of the models to reproduce these two characteristics, by simulating artificial data from the models. The MMHP models generally outperform other models, although the simpler HP model is able to capture the properties of variance jumps.

Our study suggests the clustering of jumps in financial markets may warrant more attention in the risk management literature. A possible area for future research would be to extend this approach to analyse risk premium using option data.

A Appendices for Chapter 2

A.1 Bayesian MCMC Algorithm and Specification of Priors

In order to obtain the posterior distribution in Equation (2.21), we randomly sample from a set of conditional posterior derived from Bayes's rule. A MCMC algorithm is constructed to approximate the posterior implied by model estimated. Therefore, for $i = 1, 2, ..., n, ..., N$:

1. Sample static parameters

   Draw $\Theta_1^{(i)}$ from $p\left(\Theta_1^{(i)}|Y, \Theta_2^{(i-1)}, \Theta_3^{(i-1)}, ..., \Theta_k^{(i-1)}, \xi^{P+(i-1)}, \xi^{P-(i-1)}, \xi^{V(i-1)}, \boldsymbol{\vartheta}^{(i-1)}, V^{(i-1)}\right)$,

   $\vdots$

   Draw $\Theta_k^{(i)}$ from $p\left(\Theta_k^{(i)}|Y, \Theta_1^{(i-1)}, \Theta_2^{(i-1)}, ..., \Theta_{k-1}^{(i-1)}, \xi^{P+(i-1)}, \xi^{P-(i-1)}, \xi^{V(i-1)}, \boldsymbol{\vartheta}^{(i-1)}, V^{(i-1)}\right)$

   .

2. Sample jump sizes

   for t = 1,2, ... , T:

   Draw $\xi^{P+(i)}$ from $p\left(\xi^{P+(i)}|Y, \Theta^{(i)}, \xi^{P-(i-1)}, \xi^{V(i-1)}, \boldsymbol{\vartheta}^{(i-1)}, V^{(i-1)}\right)$,

   Draw $\xi^{P-(i)}$ from $p\left(\xi^{P-(i)}|Y, \Theta^{(i)}, \xi^{P+(i)}, \xi^{V(i-1)}, \boldsymbol{\vartheta}^{(i-1)}, V^{(i-1)}\right)$,

   Draw $\xi^{V(i)}$ from $p\left(\xi^{V(i)}|Y, \Theta^{(i)}, \xi^{P+(i)}, \xi^{P-(i-1)}, \boldsymbol{\vartheta}^{(i-1)}, V^{(i-1)}\right)$.

3. Sample variance

   for t = 1,2, ... , T:

   Draw V_t from $p\left(V^{(i)}|Y, \Theta^{(i)}, \xi^{P+(i)}, \xi^{P-(i)}, \xi^{V(i)}, \boldsymbol{\vartheta}^{(i-1)}\right)$,

4. Sample Branching Coefficient Matrix

   for t = 1,2, ... , T:

   Draw $\boldsymbol{\vartheta}^{(i)}$ from $p\left(\boldsymbol{\vartheta}^{(i)}|Y, \Theta^{(i)}, \xi^{P+(i)}, \xi^{P-(i)}, \xi^{V(i)}, V^{(i)}\right)$

We set $N = 100,000$ as the total number of iterations and $n = 30,000$ as the burn-in period, which will be discarded. For those conditional posterior distributions where corresponding conjugate priors can be found and posteriors can be obtained in closed form, we adopt Gibbs sampling; for those posteriors that are unknown, we use Metropolis-Hastings (MH) to approximate posteriors. MH involves drawing a sample from a proposal density and another random number from a uniform distribution to decide whether the proposal draw should be accepted or rejected. Ultimately, we specify priors in our model as follows:

[INSERT TABLE 2.7 ABOUT HERE]

Moreover, when we run the MCMC algorithm, the original data and estimated data are amplified by 10,000 times. Results reported in Table 2.3 are transformed back and annualised, see Broadie et al. (2007) for details of this transformation technique.

A.2 Marginal Likelihood of Models

We compute marginal likelihood of data given different models as follow:

$$p(Y_t|\mathcal{M}_i) = \frac{p(Y_t|\Theta_i, \mathcal{M}_i)p(\Theta_i|\mathcal{M}_i)}{p(\Theta_i|Y_t, \mathcal{M}_i)}, \tag{IA.2.23}$$

where $i = 1, ..., 7$ denotes 7 different models, $\mathcal{M}_i$ denotes corresponding static parameters. The likelihood of data given models and static parameters is further

marginalised over k latent variables $X_t^{(k)}$ as follows:

$$p(Y_t|\Theta_i, \mathcal{M}_i) = \int p\left(Y_t|X_t^{(k)}, \Theta_i, \mathcal{M}_i\right) p\left(X_t^{(k)}|\Theta_i, \mathcal{M}_i\right) dX_t^{(k)}, \tag{IA.2.24}$$

Using the output of previous MCMC outputs, it is simply the marginal likelihood of data averaged over latent variables. Similarly, the conditional posterior of static parameters is also marginalised over latent variables:

$$p(\Theta_i|Y_t, \mathcal{M}_i) = \int p\left(\Theta_i|Y_t, X_t^{(k)}, \mathcal{M}_i\right) dX_t^{(k)}, \tag{IA.2.25}$$

Following methods proposed by Chib (1995) and Chib & Jeliazkov (2001), we decompose the static parameter vector Θ_i into two components: $\theta_{1,i}$ denoting parameters in price and variance processes, and $\theta_{2,i}$, denoting parameters in jumps components (Hawkes kernel). Therefore,

$$p(\Theta_i|Y_t, \mathcal{M}_i) = p(\theta_{1,i}|Y_t, \mathcal{M}_i)p(\theta_{2,i}|\theta_{1,i}, Y_t, \mathcal{M}_i), \tag{IA.2.26}$$

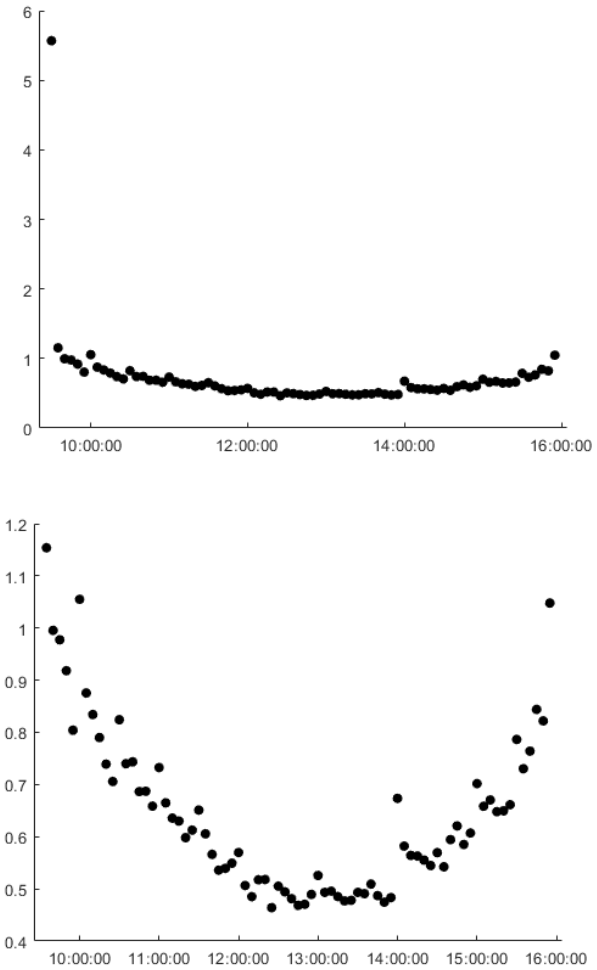
The likelihood of a multivariate HP is derived in Liniger (2009).

A.3 Full Parameter Posteriors Results

Table 2.10 to 2.21 provide full posteriors of parameters from $\mathcal{M}_1$ to $\mathcal{M}_7$ specified in the main text. These are corresponding to the Table 2.4 in the main text where we summarise these parameter results. Note, we skip results of parameters in price and variance process (Equation 2.7 and 2.8) and focus on parameters in the Hawkes kernel only. Additionally, data applied include 4 individual stock data (from 2012 to 2019), S&P 500 index data and 3 ETF data (from 2000 to 2017).

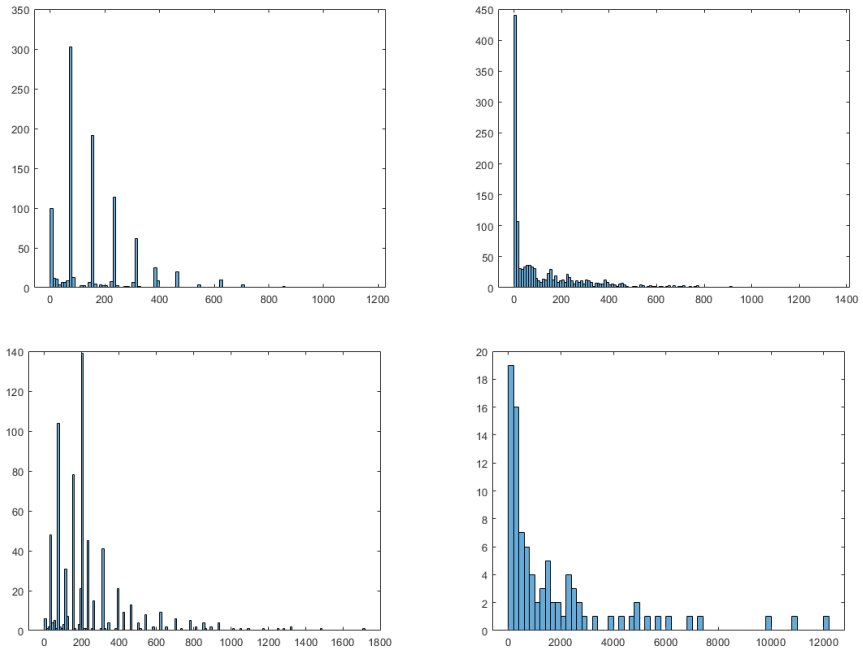
2.7 Figures and Tables for Chapter 2

Figure 2.1: Intraday Periodicity Component of S&P 500 with and without Overnight Return



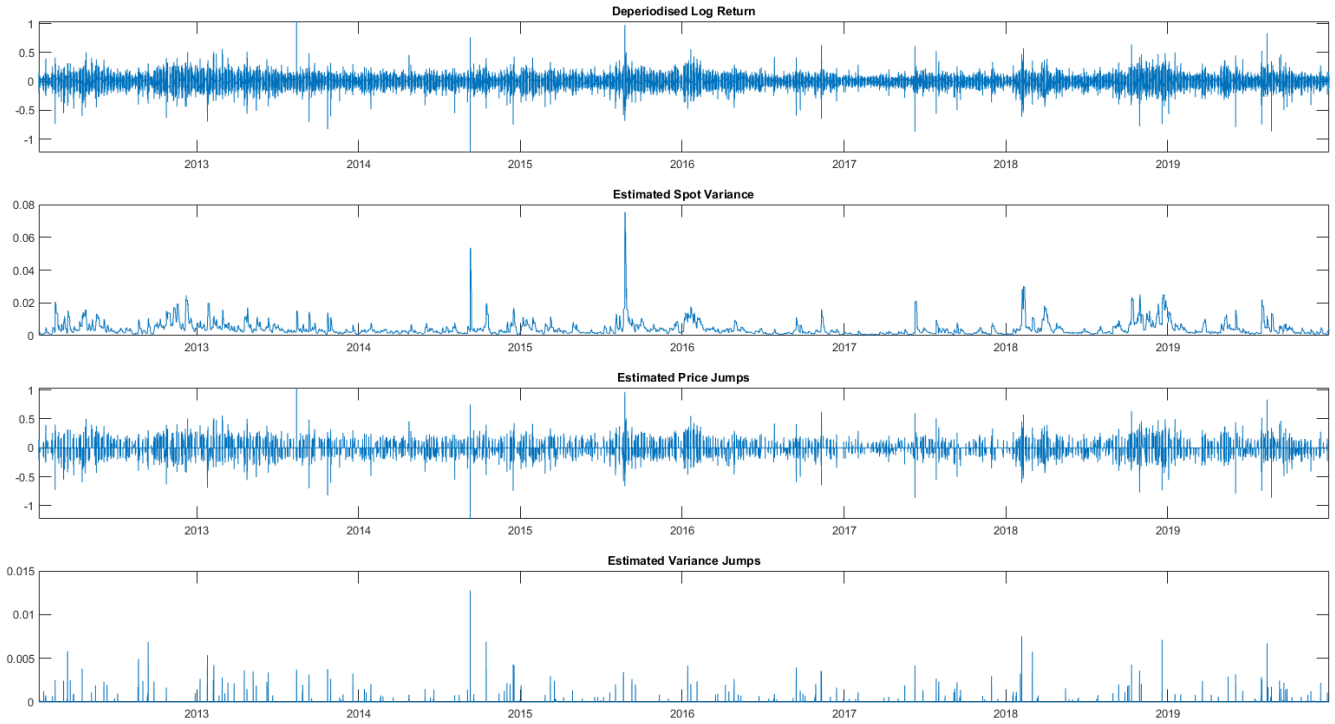
Notes: The intraday periodicity component s_t is estimated by a weighted standard deviation estimator, which is estimated using the data from 1/1/2012 to 31/12/2019

Figure 2.2: Histogram of Interarrival Times between S&P 500 Index Price(up two) and variance(down two) Jumps before(left two) and after(right two) De-periodisation (Adjusted by Intraday periodicity)



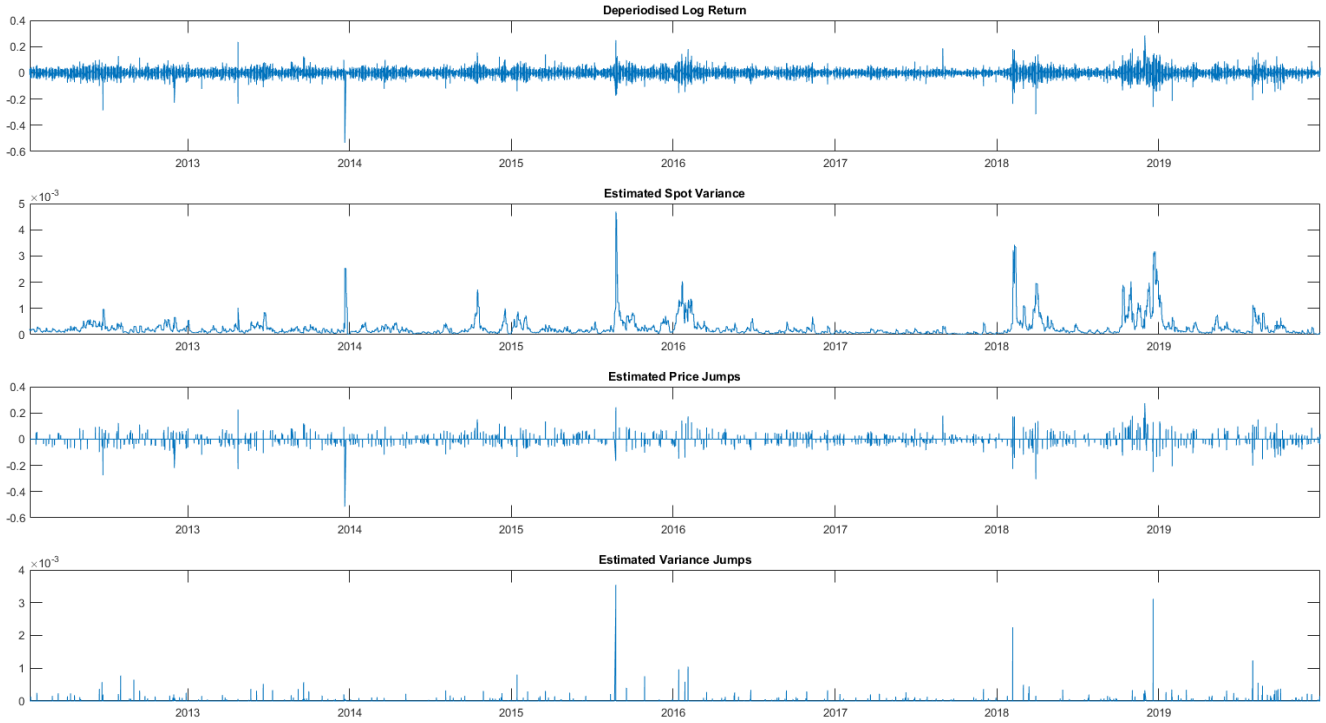
Notes: The interarrival times are calculated using the data from 1/1/2012 to 31/12/2019

Figure 2.3: Log Return, Estimated Variance, Estimated Price Jump Size and Estimated Variance Jump Size of AAPL Stock



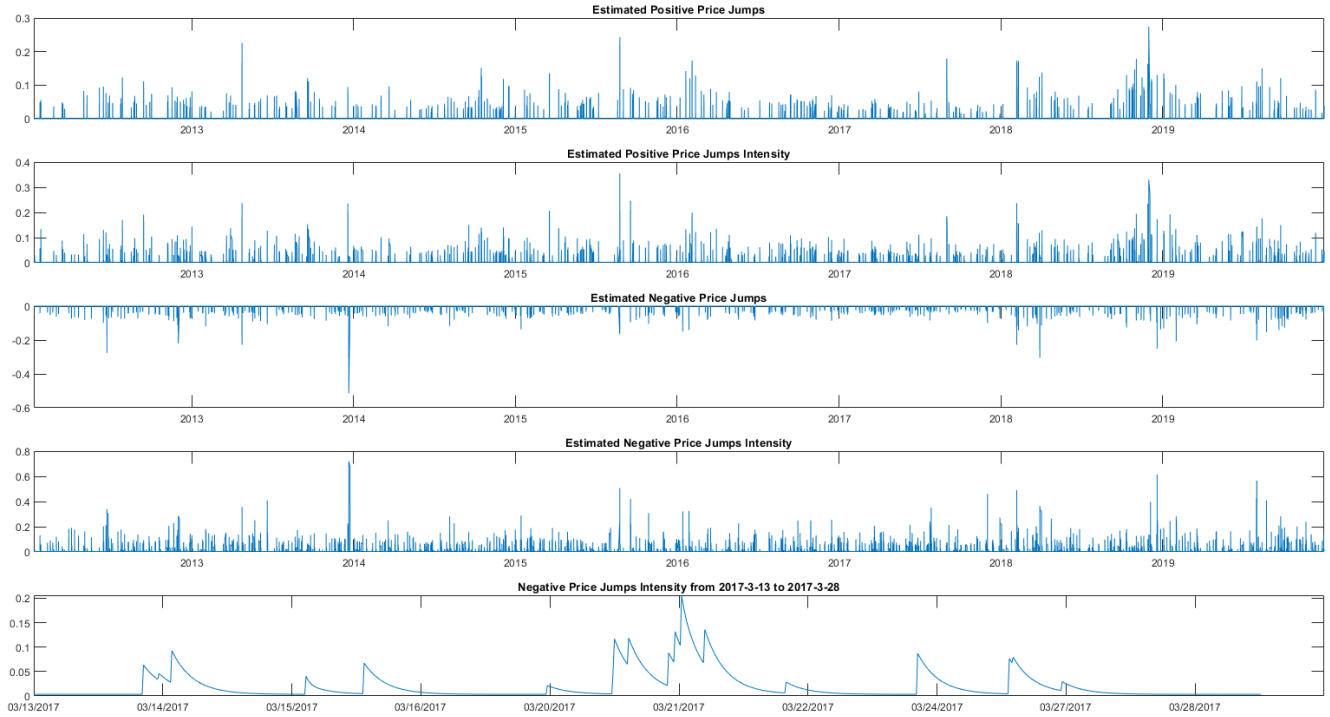
Notes: The data period is from 1/1/2012 to 31/12/2019

Figure 2.4: Log Return, Estimated Variance, Estimated Price Jump Size and Estimated Volatility Jump Size of S&P 500 Index



Notes: The data period is from 1/1/2012 to 31/12/2019

Figure 2.5: Return Jumps and Corresponding Intensities of S&P 500 Index



Notes: This figure plot positive and negative return jumps and their intensities of S&P 500 Index from 1/1/2012 to 31/12/2019. Negative jump intensity from 3/13/2017 to 3/28/2017 is plotted at the bottom. There are totally 6 jumps on 03/21/2017 when S&P 500 Index fluctuated in a range of 3% and closed at dropping 1.2%, it recorded the biggest drop since 6 months ago at the time.

Table 2.1: Number of Price and Variance Jumps in four Individual Stocks and S&P 500 Index (2012 - 2019)

No. of	AAPL (Apple Inc.)	BA (Boeing Co.)	JPM (JPMorgan Chase & Co.)	KO (Coca-Cola Co.)	SPX (S&P 500)
	Without Intraday periodicity				
Positive Price Jump	676 (0.43%)	708 (0.45%)	649 (0.41%)	406 (0.25%)	573 (0.35%)
Negative Price Jump	518 (0.33%)	583 (0.37%)	548 (0.35%)	449 (0.28%)	525 (0.32%)
Variance Jump	807 (0.51%)	885 (0.56%)	873 (0.55%)	645 (0.41%)	975 (0.59%)
	With Intraday periodicity				
Positive Price Jump	1851 (1.18%)	1519 (0.97%)	817 (0.52%)	639 (0.4%)	477 (0.29%)
Negative Price Jump	2189 (1.39%)	1451 (0.92%)	876 (0.55%)	605 (0.38%)	718 (0.44%)
Variance Jump	293 (0.18%)	270 (0.17%)	148 (0.09%)	118 (0.07%)	281 (0.17%)

Notes: Total number of observations is around 163,000.

Table 2.2: Specification of $\mathcal{M}_1$ to $\mathcal{M}_7$

	Price and variance process with jumps modelled by:
$\mathcal{M}_1$	Correlated jump size components (SVCJ) model
$\mathcal{M}_2$	Hawkes Process Without Intraday periodicity ($S_t \equiv 1$)
$\mathcal{M}_3$	Hawkes Process With Intraday periodicity
$\mathcal{M}_4$	Marked Hawkes Process with impact function (2.6, I)
$\mathcal{M}_5$	Marked Hawkes Process with impact function (2.6, II)
$\mathcal{M}_6$	Marked Hawkes Process with impact function (2.6, III)
$\mathcal{M}_7$	Marked Hawkes Process with impact function (2.6, IV)

Table 2.3: Posterior Mean and Standard Deviation of Parameters in Price and Variance Process

	AAPL		BA		JPM		KO		SPX	
	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.
$\mu(*10^{-5})$	2.819	3.056	**11.371**	3.618	-0.721	2.049	-0.264	0.655	**6.068**	0.979
γ	**0.032**	0.002	**0.042**	0.003	**0.046**	0.001	**0.119**	0.001	**0.080**	0.002
$\theta(*10^{-5})$	**51.021**	1.385	**43.744**	1.283	**38.647**	0.490	**8.392**	0.073	**3.335**	0.042
$\sigma_V(*10^{-3})$	**4.579**	0.536	**4.747**	0.936	**4.506**	0.431	**3.272**	0.302	**2.271**	0.206
ρ	**-0.150**	0.049	**-0.090**	0.047	**-0.145**	0.048	**-0.118**	0.047	**-0.131**	0.047
$\mu_Y(*10^{-3})$	-0.171	1.253	0.034	1.239	-0.009	1.492	0.013	1.492	-0.022	0.946
σ_Y	**0.2455**	0.0204	**0.2399**	0.0207	**0.3452**	0.0399	**0.3479**	0.0455	**0.1445**	0.0122
μ_V	**0.3008**	0.0246	**0.3251**	0.0277	**0.5780**	0.0660	**0.7160**	0.0914	**0.3117**	0.0261
$\xi^{P+}(*10^{-3})$	2.03	3.40	2.02	2.61	2.43	2.47	2.27	2.30	1.87	1.25
$\xi^{P-}(*10^{-3})$	-2.13	2.94	-2.09	2.98	-2.52	3.00	-2.62	4.33	-1.87	1.74
$\xi^{V}(*10^{-3})$	1.14	1.25	1.21	1.04	1.36	1.02	1.44	1.13	1.07	0.31

Notes: Parameters are estimated with data from 1/1/2012 to 31/12/2019 inflated by 78*100 times. Significant values are marked bold (confidence level: 95%).

Table 2.4: Summary of Select Parameter Posteriors

	Individual Stocks			**S&P 500**			**ETF**		
	$p+$	$p-$	v	$p+$	$p-$	v	$p+$	$p-$	v
					ϑ in $\mathcal{M}_2$				
$p+$	0.029 (4/4)	0.019 (4/4)	0.004 (0/4)	0.002 (0/1)	0.003 (0/1)	0.002 (0/1)	0.029 (1/3)	0.014 (1/3)	0.008 (1/3)
$p-$	0.04 (4/4)	0.042 (4/4)	0.006 (0/4)	0.006 (0/1)	0.007 (0/1)	0.002 (0/1)	0.019 (2/3)	0.047 (2/3)	0.012 (1/3)
v	0.014 (0/4)	0.007 (0/4)	0.003 (0/4)	0.003 (0/1)	0.002 (0/1)	0.002 (0/1)	0.004 (0/3)	0.005 (0/3)	0.014 (2/3)
					ϑ in $\mathcal{M}_3$				
$p+$	0.159 (4/4)	0.108 (4/4)	0.009 (3/4)	0.101 (1/1)	0.045 (1/1)	0.026 (1/1)	0.101 (3/3)	0.052 (3/3)	0.012 (3/3)
$p-$	0.172 (4/4)	0.216 (4/4)	0.009 (3/4)	0.116 (1/1)	0.254 (1/1)	0.009 (1/1)	0.065 (3/3)	0.191 (3/3)	0.023 (3/3)
v	0.54 (4/4)	0.409 (4/4)	0.106 (3/4)	0.265 (1/1)	0.13 (1/1)	0.04 (0/1)	0.155 (3/3)	0.103 (2/3)	0.056 (2/3)
					$\tilde{\beta}$ in $\mathcal{M}_4$				
$p+$	146.2 (3/4)	106.7 (1/4)	175.4 (0/4)	108.8 (1/1)	279.6 (0/1)	69.5 (0/0)	118.5 (2/3)	388 (3/3)	157.9 (1/3)
$p-$	152.2 (4/4)	97.5 (4/4)	200.2 (0/4)	315.1 (1/1)	84.5 (1/1)	128.4 (0/0)	379.7 (3/3)	64.7 (1/3)	44.3 (2/3)
v	90.4 (4/4)	80.95 (4/4)	41.8 (1/4)	144.5 (1/1)	98.4 (1/1)	153.6 (0/0)	53.4 (1/3)	65 (1/3)	34.1 (1/3)
					$\tilde{\gamma}$ in $\mathcal{M}_4$				
$p+$	106.4 (3/4)	326.75 (0/4)	237 (0/4)	141.7 (1/1)	871.8 (0/1)	322.4 (0/1)	85.8 (1/3)	225.1 (2/3)	131.2 (2/3)
$p-$	89.4 (0/4)	160.2 (0/4)	406.3 (4/4)	155.7 (0/1)	121.4 (0/1)	217.2 (1/1)	244.3 (2/3)	115.8 (1/3)	174 (2/3)
v	316 (0/4)	400.5 (4/4)	115.2 (4/4)	204.4 (0/1)	996.4 (1/1)	456.5 (1/1)	392.5 (2/3)	164.4 (1/3)	134.7 (2/3)
					$\tilde{\beta}$ in $\mathcal{M}_7$				
$p+$	146.1 (3/4)	123.1 (1/4)	132.9 (0/4)	61.1 (1/1)	82.8 (0/1)	198.2 (0/1)	122.7 (3/3)	418.1 (2/3)	111.5 (2/3)
$p-$	180.4 (4/4)	137.9 (4/4)	150.2 (0/4)	267 (1/1)	41.5 (0/1)	87.3 (0/1)	309.9 (3/3)	66.1 (2/3)	41.6 (2/3)
v	122.1 (0/4)	77.6 (0/4)	36.8 (4/4)	139.1 (0/1)	124.2 (0/1)	198.1 (0/1)	62 (1/3)	64.7 (0/3)	33 (0/3)

Notes: This table presents posterior means and the number of significant results of some parameters in models. The rows record the responses to the dimensions of variables in the columns. For example, ϑ denotes the mean number of jumps of one dimension that is produced by another dimension. So, the 0.04 (ϑ_{p-}^{p+} in the 2nd row, 1st column) means in individual stock data, there are, on average, 0.04 negative price jumps ($p-$) produced by positive price jumps ($p+$). Additionally, there are in total four individual stocks and this parameter is significant for all of them, hence, (4/4).

Table 2.5: Deviance Information Criterion (DIC) of $\mathcal{M}_1$ to $\mathcal{M}_7$

	AAPL	BA	JPM	KO	SPX
$\mathcal{M}_1$	509667	509182	509078	510329	509041
$\mathcal{M}_2$	509625	497819	487591	443632	417988
$\mathcal{M}_3$	493133	488032	484395	441840	417714
$\mathcal{M}_4$	492641	487608	484026	441473	416766
$\mathcal{M}_5$	492839	487773	484163	441613	416933
$\mathcal{M}_6$	494106	488585	484741	442036	417429
$\mathcal{M}_7$	492620	487599	484019	441458	416752

Notes: DIC of different models are calculated by using the data from 1/1/2012 to 31/12/2019.

Table 2.6: Log Bayes Factor (BF) of $\mathcal{M}_1$ to $\mathcal{M}_7$

	AAPL					
	$\mathcal{M}_7$	$\mathcal{M}_6$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_3$	$\mathcal{M}_2$
$\mathcal{M}_1$	8376.1	6779.1	8319.6	8368.6	7232.4	336.7
$\mathcal{M}_2$	8039.4	6442.4	7982.9	8031.9	6895.7	
$\mathcal{M}_3$	1143.7	-453.3	1087.1	1136.2		
$\mathcal{M}_4$	3.5	-1589.4	-49.0			
$\mathcal{M}_5$	56.6	-1540.4				
$\mathcal{M}_6$	1597.0					

	KO					
	$\mathcal{M}_7$	$\mathcal{M}_6$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_3$	$\mathcal{M}_2$
$\mathcal{M}_1$	1176.9	1033.3	1146.9	1167.2	1042.6	143.1
$\mathcal{M}_2$	1033.8	890.2	1003.8	1024.1	899.5	
$\mathcal{M}_3$	134.3	-9.3	104.3	124.7		
$\mathcal{M}_4$	9.7	-133.9	-20.3			
$\mathcal{M}_5$	30.0	-113.6				
$\mathcal{M}_6$	143.6					

	BA					
	$\mathcal{M}_7$	$\mathcal{M}_6$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_3$	$\mathcal{M}_2$
$\mathcal{M}_1$	5008.8	4642.8	4968.2	5001.0	4848.5	254.3
$\mathcal{M}_2$	4754.6	4388.6	4713.9	4746.7	4594.2	
$\mathcal{M}_3$	160.4	-205.6	119.7	152.5		
$\mathcal{M}_4$	4.9	-358.1	-32.8			
$\mathcal{M}_5$	40.7	-325.3				
$\mathcal{M}_6$	366.0					

	SPX					
	$\mathcal{M}_7$	$\mathcal{M}_6$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_3$	$\mathcal{M}_2$
$\mathcal{M}_1$	470.9	277.2	430.2	463.0	307.3	88.1
$\mathcal{M}_2$	382.8	189.2	342.1	374.9	219.2	
$\mathcal{M}_3$	163.6	-30.0	122.9	155.7		
$\mathcal{M}_4$	-1.9	-185.8	-32.8			
$\mathcal{M}_5$	40.7	-152.9				
$\mathcal{M}_6$	193.7					

	JPM					
	$\mathcal{M}_7$	$\mathcal{M}_6$	$\mathcal{M}_5$	$\mathcal{M}_4$	$\mathcal{M}_3$	$\mathcal{M}_2$
$\mathcal{M}_1$	1651.7	1435.9	1628.5	1646.9	1503.7	204.1
$\mathcal{M}_2$	1447.6	1231.8	1424.4	1442.8	1299.7	
$\mathcal{M}_3$	147.9	-67.8	124.7	143.1		
$\mathcal{M}_4$	4.8	-211.0	-18.4			
$\mathcal{M}_5$	23.2	-192.6				
$\mathcal{M}_6$	215.8					

Notes: The table presents the log Bayes factor of $\mathcal{M}_7$ to $\mathcal{M}_2$ (row) against $\mathcal{M}_1$ to $\mathcal{M}_6$ (column). They are calculated by using the data from 1/1/2012 to 31/12/2019.

Table 2.7: Priors Specification

General Parameters		$\mathcal{M}_4$**'s Impact Function**	
μ	$N(0,1)$	$\tilde{\alpha}$	$N(0,0.2)\mathbf{1}_{\tilde{\alpha}>0}$
ξ	$N(0,1)\mathbf{1}_{\xi>0}$	$\tilde{\beta}$	$N(0,10)\mathbf{1}_{\tilde{\beta}>0}$
θ	$N(0,1)\mathbf{1}_{\theta>0}$	$\tilde{\gamma}$	$N(0,30)\mathbf{1}_{\tilde{\gamma}>0}$
σ_V	$IG(2.5,0.1)$		
ρ	$U(-1,1)$	$\mathcal{M}_5$**'s Impact Function**	
ξ^{P+}	$N(0,50)\mathbf{1}_{\xi^{P+}>0}$	$\tilde{\alpha}$	$N(0,0.3)\mathbf{1}_{\tilde{\alpha}>0}$
ξ^{P-}	$N(0,50)\mathbf{1}_{\xi^{P-}<0}$		
ξ^{V}	$N(0,10)\mathbf{1}_{\xi^{V}>0}$	$\mathcal{M}_6$**'s Impact Function**	
$\boldsymbol{\vartheta}$	$N(0,0.1)\mathbf{1}_{\boldsymbol{\vartheta}>0}$	$\tilde{\alpha}$	$N(0,30)\mathbf{1}_{\tilde{\alpha}>0}$
λ_0^{p+}	$N(0,0.001)\mathbf{1}_{\lambda_0^{p+}>0}$		
λ_0^{p-}	$N(0,0.001)\mathbf{1}_{\lambda_0^{p-}>0}$	$\mathcal{M}_7$**'s Impact Function**	
λ_0^{v}	$N(0,0.001)\mathbf{1}_{\lambda_0^{v}>0}$	$\tilde{\alpha}$	$N(0,0.2)\mathbf{1}_{\tilde{\alpha}>0}$
β	$N(0,0.3)\mathbf{1}_{\beta>0}$	$\tilde{\beta}$	$N(0,10)\mathbf{1}_{\tilde{\beta}>0}$

Notes: The table presents priors settings in MCMC algorithm of parameter estimations

Table 2.8: Summary Statistics of Clusters

		No. of cluster	M_{NJ}	SD_{NJ}	Max(NJ)	M_L	Max(L)	No. of cluster containing only 1 jump
AAPL	P^+	621	2.981	2.489	21	39	293	206
	P^-	643	3.404	2.921	24	67	395	208
	V	198	1.480	0.900	6	12	53	138
BA	P^+	607	2.502	2.038	14	37	304	247
	P^-	580	2.502	2.141	22	69	408	239
	V	213	1.268	0.574	4	11	55	168
JPM	P^+	454	1.800	1.242	9	25	274	257
	P^-	468	1.872	1.381	12	36	178	256
	V	117	1.265	0.621	4	10	31	95
KO	P^+	409	1.562	0.991	7	14	105	272
	P^-	375	1.625	1.188	10	32	263	239
	V	99	1.192	0.444	3	18	47	82
SPX	P^+	336	1.420	0.846	7	27	109	242
	P^-	386	1.860	1.215	9	39	122	203
	V	223	1.260	0.589	4	9	48	179

Notes: M_{NJ} denotes mean number of jumps in clusters, SD_{NJ} denotes standard deviation of number of jumps in clusters, max(NJ) denotes maximum number of jumps in a cluster. M_L denotes mean of cluster length (length of a cluster of jumps cover) and $Max(L)$ denotes maximum value of cluster length. Cluster length figures are rounded to integers.

Table 2.9: Simulation Results

Characteristics	**Positive Price Jump J^{P+}**					**Negative Price Jump J^{P-}**					**Variance Jump J^V**				
AAPL	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$
M_{NJ}	2.981	1.097**	1.481**	3.27	2.72	3.404	1.134**	1.571**	3.631	3.31	1.48	1.016**	1.077**	2.125**	1.678
SD_{NJ}	2.489	0.324**	0.972**	2.874	2.315	2.921	0.389**	1.105**	3.257	2.84	0.9	0.116**	0.3**	2.116**	1.312
BA	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$
M_{NJ}	2.502	1.095**	1.356**	2.862	2.592	2.502	1.096**	1.386**	2.702	2.47	1.368	1.02**	1.164†	1.743**	1.405
SD_{NJ}	2.038	0.321**	0.763**	2.397	2.117	2.141	0.321**	0.838**	2.269	1.92	0.774	0.103**	0.435*	1.399**	0.838
JPM	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$
M_{NJ}	1.8	1.052**	1.24**	2.04	1.858	1.872	1.058**	1.308**	2.271	1.875	1.265	1.011**	1.093	1.597*	1.363
SD_{NJ}	1.242	0.231**	0.592**	1.384	1.255	1.381	0.242**	0.702**	1.774	1.588	0.621	0.135*	0.381	1.013	0.714
KO	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$
M_{NJ}	1.562	1.046**	1.208**	1.773	1.595	1.632	1.043**	1.22**	1.927	1.797	1.192	1.007**	1.077	1.175	1.152
SD_{NJ}	0.991	0.215**	0.532**	1.403†	1.241	1.118	0.207**	0.555**	1.404	1.311	0.444	0.134†	0.325	0.297	0.349
SPX	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$	Data	$\mathcal{M}_1$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_7$
M_{NJ}	1.42	1.03**	1.171**	1.799*	1.526	1.86	1.051**	1.347**	2.148	1.749	1.26	1.021**	1.042**	2.241**	1.477
SD_{NJ}	0.846	0.174**	0.461**	1.327**	1.011	1.215	0.226**	0.763**	1.637*	1.267	0.589	0.1**	0.251	2.291**	1.127†

Notes: M_{NJ} denotes the mean of the number of jumps in clusters, and SD_{NJ} denotes standard deviation of number of jumps in clusters.
† indicates less than 10 p.c. of simulations were further out in the tails than the sample estimate.
* indicates less than 5 p.c. of simulations were further out in the tails than the sample estimate.
** indicates less than 1 p.c. of simulations were further out in the tails than the sample estimate.

Table 2.10: Posterior Mean and Standard Deviation of Parameters in a 3-dimensional Hawkes Structure Model (without impact of marks and before de-periodisation)

	AAPL		BA		JPM		KO		SPX	
	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.
λ_0^{p+}	**0.00330**	0.00014	**0.00331**	0.00014	**0.00352**	0.00015	**0.00241**	0.00013	**0.00355**	0.00015
λ_0^{p-}	**0.00223**	0.00012	**0.00277**	0.00013	**0.00304**	0.00014	**0.00262**	0.00013	**0.00327**	0.00014
λ_0^{v}	**0.00342**	0.00014	**0.00382**	0.00015	**0.00369**	0.00015	**0.00266**	0.00013	**0.00415**	0.00016
ϑ_{p+}^{p+}	**0.03554**	0.01182	**0.03620**	0.01210	**0.01960**	0.00947	**0.02536**	0.00881	0.00221	0.00198
ϑ_{p-}^{p+}	**0.04232**	0.01299	**0.04612**	0.01358	**0.02118**	0.00900	**0.04953**	0.01255	0.00633	0.00497
ϑ_{v}^{p+}	0.01008	0.00722	0.02093	0.01226	0.01524	0.00894	0.00943	0.00716	0.00252	0.00217
ϑ_{p+}^{p-}	**0.01497**	0.00708	**0.01453**	0.00712	**0.01334**	0.00630	**0.03505**	0.01175	0.00267	0.00223
ϑ_{p-}^{p-}	**0.04958**	0.01412	**0.04782**	0.01516	**0.00990**	0.00474	**0.05925**	0.01444	0.00728	0.00494
ϑ_{v}^{p-}	0.00409	0.00368	0.01232	0.00824	0.00479	0.00383	0.00537	0.00478	0.00166	0.00153
ϑ_{p+}^{v}	0.00240	0.00221	0.00223	0.00205	0.00183	0.00167	0.00755	0.00591	0.00185	0.00168
ϑ_{p-}^{v}	0.01002	0.00610	0.00677	0.00458	0.00199	0.00186	0.00613	0.00428	0.00164	0.00156
ϑ_{v}^{v}	0.00191	0.00181	0.00252	0.00228	0.00167	0.00157	0.00585	0.00489	0.00160	0.00150
β_{p+}^{p+}	**0.68178**	0.16344	**0.59151**	0.12052	**0.57802**	0.14065	**0.76889**	0.18061	**0.74664**	0.19683
β_{p-}^{p+}	**0.56162**	0.10462	**0.60227**	0.09931	**0.54975**	0.10842	**0.46517**	0.07440	**0.50810**	0.19709
β_{v}^{p+}	**0.81568**	0.18226	**0.67359**	0.14879	**0.68958**	0.15207	**0.69433**	0.17980	**0.91608**	0.19418
β_{p+}^{p-}	**0.71016**	0.20938	**0.63989**	0.18798	**0.68816**	0.21710	**0.81714**	0.20306	**1.01213**	0.31245
β_{p-}^{p-}	**0.50305**	0.10923	**0.45823**	0.08488	**0.82526**	0.21626	**0.39147**	0.06942	**0.63878**	0.20944
β_{v}^{p-}	**1.32656**	0.20856	**0.76492**	0.23168	**1.14300**	0.23922	**1.35068**	0.20559	**1.62522**	0.20297
β_{p+}^{v}	**0.81999**	0.38924	**0.84159**	0.39234	**0.82782**	0.39647	**0.91125**	0.36048	**0.80049**	0.39342
β_{p-}^{v}	**1.17967**	0.32656	**1.12093**	0.38056	**1.14203**	0.51008	**1.12435**	0.42582	**1.13095**	0.52980
β_{v}^{v}	**0.86550**	0.35562	**0.88270**	0.33751	**0.87288**	0.36475	**0.87369**	0.31989	**0.90542**	0.36437

Notes: The posterior mean and standard deviation of parameters presented here are from Hawkes Process kernel without impact of marks and before the de-periodisation of asset returns, λ_0 denotes ground intensity, ϑ_q^i denotes average number jumps in dimension i directly induced by jumps in dimension q and β_q^i denotes corresponding decaying speeds. Values that are significant greater than 0 are marked bold.

Table 2.11: Posterior Mean and Standard Deviation of Parameters in a 3-dimensional Hawkes Structure Model (without impact of marks and after de-periodisation)

	AAPL		BA		JPM		KO		SPX	
	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.
λ_0^{p+}	**0.00614**	0.00025	**0.00605**	0.00022	**0.00351**	0.00017	**0.00294**	0.00015	**0.00224**	0.00013
λ_0^{p-}	**0.00832**	0.00026	**0.00587**	0.00022	**0.00378**	0.00017	**0.00287**	0.00015	**0.00354**	0.00016
λ_0^{v}	**0.00032**	0.00005	**0.00037**	0.00006	**0.00021**	0.00004	**0.00013**	0.00003	**0.00045**	0.00006
ϑ_{p+}^{p+}	**0.20806**	0.01786	**0.16167**	0.01974	**0.13636**	0.02342	**0.13007**	0.02281	**0.10092**	0.02319
ϑ_{p-}^{p+}	**0.20664**	0.01653	**0.19215**	0.02182	**0.14830**	0.02288	**0.14247**	0.02651	**0.11623**	0.02015
ϑ_{v}^{p+}	**0.51729**	0.03704	**0.52772**	0.11636	**0.63656**	0.14956	**0.47720**	0.17511	**0.26506**	0.07930
ϑ_{p+}^{p-}	**0.13529**	0.01650	**0.10573**	0.01740	**0.10586**	0.02271	**0.08574**	0.02524	**0.04465**	0.01640
ϑ_{p-}^{p-}	**0.25920**	0.01818	**0.23425**	0.02314	**0.20301**	0.02422	**0.16834**	0.02646	**0.25364**	0.02764
ϑ_{v}^{p-}	**0.69279**	0.06351	**0.39147**	0.07927	**0.47086**	0.12745	**0.07905**	0.03002	**0.12979**	0.04001
ϑ_{p+}^{v}	**0.01044**	0.00324	**0.00716**	0.00319	0.00596	0.00329	**0.01071**	0.00511	**0.02611**	0.00963
ϑ_{p-}^{v}	**0.01344**	0.00296	**0.00972**	0.00367	**0.00714**	0.00305	0.00395	0.00258	**0.00923**	0.00417
ϑ_{v}^{v}	**0.07827**	0.00654	**0.17261**	0.04368	**0.09809**	0.04709	0.07522	0.04262	0.03961	0.02038
β_{p+}^{p+}	**0.12117**	0.01041	**0.11566**	0.01547	**0.09937**	0.01811	**0.12295**	0.02234	**0.21823**	0.04982
β_{p-}^{p+}	**0.11220**	0.00999	**0.10362**	0.01261	**0.10945**	0.01710	**0.09344**	0.01964	**0.08306**	0.01326
β_{v}^{p+}	**0.20506**	0.01952	**0.30404**	0.05798	**0.31869**	0.05943	**0.25182**	0.06742	**0.20584**	0.04950
β_{p+}^{p-}	**0.08885**	0.01174	**0.08602**	0.01493	**0.09502**	0.02363	**0.14859**	0.06176	**0.10027**	0.04142
β_{p-}^{p-}	**0.11870**	0.00848	**0.07749**	0.00767	**0.10688**	0.01323	**0.10233**	0.01681	**0.12371**	0.01361
β_{v}^{p-}	**0.20236**	0.02960	**0.20768**	0.04413	**0.12934**	0.03286	**0.80471**	0.26233	**0.50963**	0.13780
β_{p+}^{v}	**0.22734**	0.04973	**0.50029**	0.20026	**0.60695**	0.29642	**0.77601**	0.29353	**0.39222**	0.13254
β_{p-}^{v}	**0.58238**	0.07723	**0.28513**	0.09334	**0.75154**	0.23179	**0.98805**	0.39534	**0.73104**	0.26899
β_{v}^{v}	**0.81287**	0.08481	**0.45652**	0.09065	**0.37431**	0.17468	**0.58253**	0.26971	**0.84221**	0.28391

Notes: The posterior mean and standard deviation of parameters presented here are from Hawkes Process kernel without impact of marks and after the de-periodisation of asset returns, λ_0 denotes ground intensity, ϑ_q^i denotes average number jumps in dimension i directly induced by jumps in dimension q and β_q^i denotes corresponding decaying speeds. Values that are significant greater than 0 are marked bold.

Table 2.12: Posterior Mean and Standard Deviation of Parameters of Marked Hawkes kernel in $\mathcal{M}_4$ with Impact Function in (3,I)

	AAPL		BA		JPM		KO		SPX	
	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.
λ_0^{p+}	**0.00612**	0.00023	**0.00606**	0.00023	**0.00352**	0.00017	**0.00294**	0.00015	**0.00204**	0.00012
λ_0^{p-}	**0.00831**	0.00027	**0.00585**	0.00023	**0.00378**	0.00017	**0.00285**	0.00015	**0.00316**	0.00015
λ_0^{v}	**0.00032**	0.00006	**0.00037**	0.00006	**0.00021**	0.00004	**0.00013**	0.00003	**0.00045**	0.00006
ϑ_{p+}^{p+}	**0.20829**	0.02419	**0.16158**	0.03535	**0.13168**	0.03988	**0.13486**	0.04573	**0.09832**	0.03326
ϑ_{p-}^{p+}	**0.21114**	0.03222	**0.19654**	0.04000	**0.15552**	0.04286	**0.14697**	0.04857	**0.11868**	0.03603
ϑ_{v}^{p+}	**0.45925**	0.10253	**0.43934**	0.08845	**0.51326**	0.09822	**0.20127**	0.07229	**0.15662**	0.06325
ϑ_{p+}^{p-}	**0.13627**	0.02443	**0.10744**	0.03145	**0.10404**	0.03962	**0.09896**	0.05475	**0.04209**	0.02363
ϑ_{p-}^{p-}	**0.27075**	0.02750	**0.23857**	0.04907	**0.20978**	0.05534	**0.16519**	0.04724	**0.24716**	0.05054
ϑ_{v}^{p-}	**0.63048**	0.10784	**0.31285**	0.08253	**0.37472**	0.12442	**0.03664**	0.01949	**0.08132**	0.03594
ϑ_{p+}^{v}	**0.01067**	0.00632	**0.00793**	0.00473	**0.00789**	0.00521	**0.01183**	0.00631	**0.02156**	0.00920
ϑ_{p-}^{v}	**0.01336**	0.00703	**0.00991**	0.00601	**0.00739**	0.00534	0.00490	0.00414	**0.01498**	0.00930
ϑ_{v}^{v}	**0.07778**	0.03655	**0.16434**	0.05265	**0.06754**	0.04484	0.04852	0.02975	0.03448	0.02189
$\tilde{\alpha}_{p+}^{p+}$	**0.97416**	0.09700	**0.76704**	0.13990	**0.71245**	0.18259	**0.70893**	0.19824	**0.71323**	0.23096
$\tilde{\alpha}_{p-}^{p+}$	**0.72934**	0.09873	**0.76222**	0.13167	**0.67490**	0.17297	**0.54608**	0.19875	**0.41058**	0.17879
$\tilde{\alpha}_{v}^{p+}$	**0.93715**	0.19454	**0.94516**	0.17803	**0.93956**	0.16687	**0.92654**	0.32840	**0.84887**	0.33894
$\tilde{\alpha}_{p+}^{p-}$	**0.96261**	0.14692	**0.85797**	0.19914	**0.82133**	0.25826	**0.75185**	0.42238	**0.81270**	0.40030
$\tilde{\alpha}_{p-}^{p-}$	**0.94005**	0.07901	**0.74820**	0.11945	**0.71399**	0.15097	**0.23268**	0.13182	**0.79495**	0.12775
$\tilde{\alpha}_{v}^{p-}$	**0.95978**	0.16387	**0.92208**	0.25022	**0.88269**	0.31130	**0.85018**	0.50531	**0.89902**	0.42340
$\tilde{\alpha}_{p+}^{v}$	**0.76874**	0.37434	**0.66766**	0.37879	**0.70095**	0.44143	**0.70675**	0.37005	**0.89826**	0.35017
$\tilde{\alpha}_{p-}^{v}$	**0.61114**	0.29131	**0.56857**	0.35743	**0.62784**	0.40925	**0.69567**	0.53492	**0.59252**	0.38865
$\tilde{\alpha}_{v}^{v}$	**0.88711**	0.40324	**0.91462**	0.27271	**0.75310**	0.55147	**0.84580**	0.53004	**0.81844**	0.53569
$\tilde{\beta}_{p+}^{p+}$	12.336	9.266	**114.598**	38.797	**117.469**	49.177	**127.306**	61.701	**152.884**	57.198
$\tilde{\beta}_{p-}^{p+}$	**126.667**	25.019	**113.505**	34.130	**128.460**	40.277	**172.487**	49.704	**314.183**	66.177
$\tilde{\beta}_{v}^{p+}$	**54.909**	25.092	**45.207**	19.200	**44.557**	18.055	**50.909**	21.313	**141.187**	60.648
$\tilde{\beta}_{p+}^{p-}$	17.911	15.513	69.594	45.746	72.918	49.944	108.575	57.129	99.159	85.393
$\tilde{\beta}_{p-}^{p-}$	**27.951**	10.489	**120.237**	41.126	**113.076**	44.550	**292.318**	58.596	**109.312**	40.885
$\tilde{\beta}_{v}^{p-}$	**35.135**	6.266	**64.226**	11.175	**86.451**	15.259	**103.813**	18.359	**94.327**	17.310
$\tilde{\beta}_{p+}^{v}$	112.355	105.966	163.111	106.549	121.636	88.755	128.249	71.296	53.837	40.402
$\tilde{\beta}_{p-}^{v}$	181.231	109.559	204.412	118.370	146.250	123.512	114.385	117.338	216.466	123.595
$\tilde{\beta}_{v}^{v}$	98.630	58.258	70.364	39.277	**181.977**	82.841	106.838	57.549	169.595	92.600
$\tilde{\gamma}_{p+}^{p+}$	**186.560**	67.756	314.537	115.346	**422.911**	151.209	**329.930**	118.325	**303.706**	116.684
$\tilde{\gamma}_{p-}^{p+}$	125.282	117.297	141.093	126.830	179.378	166.713	238.005	200.273	286.940	231.737
$\tilde{\gamma}_{v}^{p+}$	56.276	29.719	39.842	20.747	35.468	18.393	44.973	23.352	127.693	66.238
$\tilde{\gamma}_{p+}^{p-}$	238.799	196.609	326.542	259.990	292.087	233.692	263.461	197.922	569.203	452.882
$\tilde{\gamma}_{p-}^{p-}$	77.679	43.561	131.250	73.712	132.494	73.551	120.568	68.910	93.167	51.580
$\tilde{\gamma}_{v}^{p-}$	**39.580**	16.308	**63.365**	25.890	**87.073**	35.646	**99.123**	40.529	**90.342**	36.793
$\tilde{\gamma}_{p+}^{v}$	705.482	562.387	634.251	514.056	658.541	557.710	335.421	279.414	322.709	274.488
$\tilde{\gamma}_{p-}^{v}$	**531.400**	244.004	**990.111**	447.520	**524.398**	231.843	**640.237**	279.640	**591.601**	258.801
$\tilde{\gamma}_{v}^{v}$	**106.996**	34.276	**73.700**	23.644	**164.735**	53.434	**107.278**	35.115	**174.010**	55.825
β_{p+}^{p+}	**0.12158**	0.01149	**0.11739**	0.01631	**0.10952**	0.02008	**0.12870**	0.02246	**0.19814**	0.03905
β_{p-}^{p+}	**0.11080**	0.01077	**0.10442**	0.01343	**0.10575**	0.01727	**0.09084**	0.01949	**0.09205**	0.01440
β_{v}^{p+}	**0.22777**	0.03936	**0.34432**	0.06662	**0.33201**	0.06631	**0.66478**	0.22257	**0.28837**	0.11187
β_{p+}^{p-}	**0.08952**	0.01297	**0.08422**	0.01623	**0.09802**	0.02793	**0.10356**	0.04791	**0.11448**	0.05412
β_{p-}^{p-}	**0.12016**	0.00961	**0.07871**	0.00818	**0.10687**	0.01497	**0.12535**	0.01846	**0.13088**	0.01691
β_{v}^{p-}	**0.19048**	0.03651	**0.24269**	0.07333	**0.14678**	0.05520	**1.32189**	0.25565	**0.63149**	0.22212
β_{p+}^{v}	**0.21374**	0.06768	**0.46466**	0.11664	**0.47112**	0.13588	**0.60953**	0.14057	**0.37113**	0.09652
β_{p-}^{v}	**0.51603**	0.08436	**0.33698**	0.09702	**0.98972**	0.39125	**1.25219**	0.53831	**0.44950**	0.19617
β_{v}^{v}	**0.69185**	0.28231	**0.47511**	0.10024	**0.51419**	0.25153	**1.09478**	0.40708	**0.92367**	0.34199

Table 2.13: Posterior Mean and Standard Deviation of Parameters of Marked Hawkes kernel in $\mathcal{M}_5$ with Impact Function in (3,II)

	AAPL		BA		JPM		KO		SPX	
	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.
λ_0^{p+}	**0.00650**	0.00024	**0.00623**	0.00024	**0.00357**	0.00017	**0.00298**	0.00015	**0.00214**	0.00013
λ_0^{p-}	**0.00867**	0.00026	**0.00606**	0.00023	**0.00386**	0.00017	**0.00292**	0.00015	**0.00316**	0.00016
λ_0^{v}	**0.00033**	0.00006	**0.00037**	0.00006	**0.00022**	0.00004	**0.00013**	0.00003	**0.00046**	0.00006
ϑ_{p+}^{p+}	**0.13332**	0.01230	**0.11131**	0.01460	**0.09824**	0.01769	**0.09687**	0.01784	**0.02180**	0.00531
ϑ_{p-}^{p+}	**0.12844**	0.01256	**0.12504**	0.01537	**0.09960**	0.01574	**0.09319**	0.01745	**0.06339**	0.01163
ϑ_{v}^{p+}	**0.52265**	0.08956	**0.49529**	0.09203	**0.52104**	0.11734	**0.34839**	0.11284	**0.31874**	0.07431
ϑ_{p+}^{p-}	**0.09290**	0.01281	**0.07571**	0.01308	**0.07407**	0.01623	**0.07468**	0.01777	**0.03319**	0.01224
ϑ_{p-}^{p-}	**0.15790**	0.01246	**0.15170**	0.01622	**0.13303**	0.01698	**0.09998**	0.01765	**0.11956**	0.01460
ϑ_{v}^{p-}	**0.64707**	0.10851	**0.40574**	0.08826	**0.46309**	0.12245	**0.02573**	0.02091	**0.22188**	0.06259
ϑ_{p+}^{v}	**0.00653**	0.00217	**0.00413**	0.00162	**0.00264**	0.00146	**0.00661**	0.00242	**0.00857**	0.00258
ϑ_{p-}^{v}	**0.00824**	0.00222	**0.00621**	0.00244	**0.00438**	0.00201	0.00137	0.00088	**0.00567**	0.00259
ϑ_{v}^{v}	**0.07668**	0.02568	**0.14199**	0.03843	**0.05963**	0.02970	0.03804	0.02263	**0.03692**	0.01790
$\tilde{\alpha}_{p+}^{p+}$	**1.35343**	0.02302	**1.43771**	0.03724	**1.59853**	0.06256	**1.47903**	0.05589	**1.74059**	0.10898
$\tilde{\alpha}_{p-}^{p+}$	**1.40912**	0.02612	**1.43862**	0.03463	**1.51280**	0.04855	**1.60243**	0.06482	**1.75226**	0.08034
$\tilde{\alpha}_{v}^{p+}$	**0.68348**	0.01697	**0.60081**	0.01570	**0.56284**	0.01789	**0.75482**	0.03983	**0.85844**	0.03261
$\tilde{\alpha}_{p+}^{p-}$	**1.54341**	0.04486	**1.64435**	0.06515	**1.67849**	0.08592	**1.57839**	0.08441	**2.17473**	0.32345
$\tilde{\alpha}_{p-}^{p-}$	**1.29636**	0.01807	**1.46012**	0.03068	**1.43374**	0.03498	**1.55487**	0.05736	**1.44522**	0.03494
$\tilde{\alpha}_{v}^{p-}$	**0.69368**	0.01689	**0.79001**	0.02679	**0.87784**	0.03858	**1.11668**	0.30002	**0.77644**	0.03540
$\tilde{\alpha}_{p+}^{v}$	**2.19339**	0.28668	**2.00773**	0.27217	**2.09955**	0.47320	**1.70393**	0.16487	**1.83283**	0.15765
$\tilde{\alpha}_{p-}^{v}$	**1.82858**	0.13467	**2.16948**	0.33942	**2.06214**	0.33973	**2.46239**	0.98466	**1.93706**	0.30765
$\tilde{\alpha}_{v}^{v}$	**1.04050**	0.06671	**0.82261**	0.03644	**1.20562**	0.14069	**1.23322**	0.19042	**1.13290**	0.12418
β_{p+}^{p+}	**0.11308**	0.01042	**0.10391**	0.01632	**0.08273**	0.01650	**0.11455**	0.02208	**0.19360**	0.05507
β_{p-}^{p+}	**0.10217**	0.01120	**0.09445**	0.01368	**0.10875**	0.01945	**0.09171**	0.02168	**0.06432**	0.01300
β_{v}^{p+}	**0.18899**	0.02603	**0.26852**	0.04283	**0.29923**	0.05110	**0.24300**	0.07002	**0.16650**	0.03103
β_{p+}^{p-}	**0.09006**	0.01276	**0.08047**	0.01573	**0.08624**	0.02374	**0.10981**	0.03711	**0.03252**	0.01658
β_{p-}^{p-}	**0.11755**	0.00959	**0.07286**	0.00854	**0.10314**	0.01455	**0.09846**	0.01954	**0.08943**	0.01330
β_{v}^{p-}	**0.14747**	0.02401	**0.17366**	0.03637	**0.11899**	0.02694	**1.00748**	0.35908	**0.31662**	0.08099
β_{p+}^{v}	**0.17083**	0.06222	**0.47735**	0.12837	**0.68198**	0.17002	**0.84676**	0.16036	**0.36851**	0.10899
β_{p-}^{v}	**0.43831**	0.11316	**0.19947**	0.09075	**0.47813**	0.20911	**0.48289**	0.24102	**0.40234**	0.19356
β_{v}^{v}	**0.38443**	0.11985	**0.44479**	0.08216	**0.31307**	0.11433	**0.46535**	0.17524	**0.56217**	0.18991

Table 2.14: Posterior Mean and Standard Deviation of Parameters of Marked Hawkes kernel in $\mathcal{M}_6$ with Impact Function in (3,III)

	AAPL		BA		JPM		KO		SPX	
	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.
λ_0^{p+}	**0.01143**	0.00056	**0.00915**	0.00068	**0.00507**	0.00018	**0.00400**	0.00016	**0.00305**	0.00014
λ_0^{p-}	**0.01371**	0.00047	**0.00827**	0.00126	**0.00545**	0.00018	**0.00379**	0.00015	**0.00456**	0.00021
λ_0^{v}	**0.00060**	0.00007	**0.00056**	0.00007	**0.00028**	0.00005	**0.00017**	0.00004	**0.00060**	0.00007
ϑ_{p+}^{p+}	**0.00006**	0.00000	**0.00007**	0.00000	**0.00011**	-0.00001	**0.00009**	0.00000	**0.00005**	0.00000
ϑ_{p-}^{p+}	**0.00007**	0.00000	**0.00006**	0.00000	**0.00009**	0.00000	**0.00007**	0.00000	**0.00006**	0.00000
ϑ_{v}^{p+}	**0.00008**	-0.00002	**0.00010**	-0.00003	**0.00016**	-0.00007	**0.00013**	-0.00006	**0.00005**	0.00000
ϑ_{p+}^{p-}	**0.00006**	0.00000	**0.00006**	-0.00001	**0.00008**	-0.00001	**0.00009**	0.00000	**0.00005**	0.00000
ϑ_{p-}^{p-}	**0.00006**	0.00000	**0.00006**	0.00000	**0.00008**	0.00000	**0.00007**	0.00000	**0.00005**	0.00000
ϑ_{v}^{p-}	**0.00009**	-0.00002	**0.00009**	-0.00002	**0.00011**	-0.00004	**0.00013**	-0.00005	**0.00005**	0.00000
ϑ_{p+}^{v}	**0.00006**	0.00000	**0.00006**	-0.00001	**0.00007**	-0.00001	**0.00007**	-0.00001	**0.00005**	0.00000
ϑ_{p-}^{v}	**0.00006**	0.00000	**0.00006**	0.00000	**0.00007**	-0.00001	**0.00006**	-0.00001	**0.00006**	0.00000
ϑ_{v}^{v}	**0.00011**	-0.00002	**0.00011**	-0.00002	**0.00030**	-0.00007	**0.00026**	-0.00009	**0.00005**	0.00000
$\tilde{\alpha}_{p+}^{p+}$	**76.87**	38.39	**172.94**	14.69	**309.63**	24.90	**274.61**	21.53	33.54	24.58
$\tilde{\alpha}_{p-}^{p+}$	**120.87**	11.81	**86.75**	23.00	**221.57**	18.99	**127.52**	15.24	**76.98**	18.27
$\tilde{\alpha}_{v}^{p+}$	450.43	270.72	**544.29**	270.49	842.50	454.42	653.43	394.25	22.18	13.20
$\tilde{\alpha}_{p+}^{p-}$	**120.73**	30.95	**97.21**	46.85	**196.46**	69.19	**245.04**	22.57	**20.10**	9.97
$\tilde{\alpha}_{p-}^{p-}$	**65.85**	29.25	**86.48**	21.36	**207.23**	12.22	**107.40**	29.11	**41.31**	17.38
$\tilde{\alpha}_{v}^{p-}$	459.55	246.12	442.22	230.43	601.77	359.60	637.63	394.25	**36.18**	17.38
$\tilde{\alpha}_{p+}^{v}$	59.55	33.24	**91.62**	45.44	**158.55**	78.08	**151.82**	73.35	**59.91**	13.95
$\tilde{\alpha}_{p-}^{v}$	**61.61**	30.22	**64.34**	32.29	122.71	65.94	66.93	37.50	**114.39**	20.98
$\tilde{\alpha}_{v}^{v}$	**645.59**	211.22	**624.81**	191.26	**1313.45**	193.38	**1134.12**	293.83	**84.69**	41.04
β_{p+}^{p+}	0.04589	0.03880	0.03979	0.02294	**0.19189**	0.09594	0.08542	0.04899	0.35121	0.19313
β_{p-}^{p+}	**0.03244**	0.01647	0.01941	0.01833	**0.05512**	0.02359	0.07072	0.04668	**1.19699**	0.17788
β_{v}^{p+}	0.46667	0.28598	0.38110	0.27483	0.45879	0.27567	0.53490	0.28477	0.42752	0.26675
β_{p+}^{p-}	0.11988	0.08915	0.04028	0.03321	0.07900	0.05587	**0.02598**	0.01278	0.33085	0.21382
β_{p-}^{p-}	0.02449	0.02402	0.02443	0.02497	**0.02562**	0.00962	0.10978	0.09547	0.08839	0.09358
β_{v}^{p-}	**1.86442**	0.67745	1.26552	0.66575	**1.81690**	0.68076	**1.86605**	0.68096	**1.73658**	0.65074
β_{p+}^{v}	0.14406	0.13055	0.14933	0.12080	**0.38713**	0.17633	0.25034	0.14138	**0.51891**	0.21285
β_{p-}^{v}	0.09106	0.08775	0.18525	0.17740	0.38869	0.27276	0.39946	0.30728	0.19799	0.16801
β_{v}^{v}	0.45538	0.25960	0.34455	0.19880	**0.83013**	0.36459	**0.71708**	0.35528	**0.57692**	0.27855

Table 2.15: Posterior Mean and Standard Deviation of Parameters of Marked Hawkes kernel in $\mathcal{M}_7$ with Impact Function in (3,IV)

	AAPL		BA		JPM		KO		SPX	
	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.
λ_0^{p+}	**0.00613**	0.00024	**0.00608**	0.00023	**0.00353**	0.00016	**0.00295**	0.00015	**0.00204**	0.00012
λ_0^{p-}	**0.00832**	0.00026	**0.00586**	0.00022	**0.00380**	0.00017	**0.00285**	0.00015	**0.00316**	0.00015
λ_0^{v}	**0.00032**	0.00005	**0.00037**	0.00006	**0.00022**	0.00004	**0.00013**	0.00003	**0.00045**	0.00006
ϑ_{p+}^{p+}	**0.20862**	0.02615	**0.16109**	0.03259	**0.13096**	0.03569	**0.13186**	0.04370	**0.09918**	0.03948
ϑ_{p-}^{p+}	**0.21149**	0.03561	**0.19730**	0.04604	**0.15698**	0.04828	**0.14827**	0.05173	**0.11890**	0.04446
ϑ_{v}^{p+}	**0.44549**	0.09725	**0.42280**	0.08843	**0.49393**	0.08588	**0.27489**	0.11581	**0.14952**	0.06475
ϑ_{p+}^{p-}	**0.13659**	0.02663	**0.10816**	0.03491	**0.10088**	0.05103	**0.09973**	0.04895	0.04736	0.03001
ϑ_{p-}^{p-}	**0.27081**	0.02802	**0.23946**	0.04961	**0.21210**	0.05607	**0.16586**	0.05072	**0.24750**	0.05385
ϑ_{v}^{p-}	**0.63490**	0.11437	**0.30649**	0.07671	**0.36193**	0.13386	0.03164	0.01687	**0.08040**	0.03352
ϑ_{p+}^{v}	0.01070	0.00656	0.00785	0.00490	0.00808	0.00571	0.01197	0.00774	**0.02090**	0.01029
ϑ_{p-}^{v}	0.01361	0.00866	0.00908	0.00633	0.00802	0.00626	0.00491	0.00390	0.01645	0.01306
ϑ_{v}^{v}	0.08187	0.04178	**0.16946**	0.05311	0.07721	0.05674	0.07153	0.05545	0.03888	0.03206
$\tilde{\alpha}_{p+}^{p+}$	**0.96440**	0.09572	**0.70241**	0.13000	**0.61822**	0.16370	**0.65305**	0.19086	**0.62300**	0.22606
$\tilde{\alpha}_{p-}^{p+}$	**0.70786**	0.09902	**0.74073**	0.13653	**0.64089**	0.17826	**0.50842**	0.19726	**0.40153**	0.18887
$\tilde{\alpha}_{v}^{p+}$	**0.97440**	0.19769	**0.98004**	0.19410	**0.97931**	0.15848	**0.96860**	0.39723	**0.94272**	0.38951
$\tilde{\alpha}_{p+}^{p-}$	**0.94585**	0.14716	**0.77060**	0.19843	**0.58027**	0.25794	**0.51131**	0.25629	0.61649	0.33722
$\tilde{\alpha}_{p-}^{p-}$	**0.94127**	0.08157	**0.73462**	0.12013	**0.72950**	0.15446	**0.30713**	0.15102	**0.88875**	0.13791
$\tilde{\alpha}_{v}^{p-}$	**0.98659**	0.17102	**0.97663**	0.23496	**0.96178**	0.34537	0.96636	0.51104	**0.97328**	0.39955
$\tilde{\alpha}_{p+}^{v}$	0.73776	0.37681	0.60610	0.36778	0.59985	0.40224	0.61942	0.39351	**0.82641**	0.35318
$\tilde{\alpha}_{p-}^{v}$	0.64406	0.36564	0.62760	0.43047	0.52020	0.35895	0.60304	0.43494	0.49545	0.37966
$\tilde{\alpha}_{v}^{v}$	0.96067	0.49317	**0.97029**	0.30020	0.91373	0.70692	0.95266	0.75424	0.93144	0.79753
$\tilde{\beta}_{p+}^{p+}$	17.533	14.591	**147.356**	35.796	**157.514**	44.888	**152.813**	61.910	**201.927**	92.121
$\tilde{\beta}_{p-}^{p+}$	**137.153**	32.565	**124.215**	46.387	**142.579**	51.328	**187.722**	57.913	**319.850**	98.916
$\tilde{\beta}_{v}^{p+}$	22.409	18.033	16.484	12.423	15.277	11.371	21.806	16.711	53.592	40.740
$\tilde{\beta}_{p+}^{p-}$	26.671	23.553	113.589	61.586	173.174	102.264	**215.240**	103.316	205.412	158.799
$\tilde{\beta}_{p-}^{p-}$	**27.571**	10.278	**127.141**	41.694	**107.398**	43.634	**264.586**	59.111	59.457	42.580
$\tilde{\beta}_{v}^{p-}$	11.735	7.975	19.301	12.672	28.223	18.073	23.356	15.402	25.003	16.298
$\tilde{\beta}_{p+}^{v}$	129.163	116.298	195.043	127.331	165.095	125.401	167.624	111.260	92.976	74.566
$\tilde{\beta}_{p-}^{v}$	167.105	127.181	178.418	127.805	190.498	167.410	151.589	137.256	269.653	221.244
$\tilde{\beta}_{v}^{v}$	**34.427**	15.035	**24.529**	10.915	**63.710**	20.646	**32.870**	14.516	64.142	25.258
β_{p+}^{p+}	**0.12229**	0.01143	**0.12133**	0.01628	**0.11723**	0.02051	**0.13274**	0.02188	**0.20435**	0.03835
β_{p-}^{p+}	**0.11111**	0.01070	**0.10356**	0.01330	**0.10446**	0.01699	**0.09084**	0.01915	**0.09217**	0.01469
β_{v}^{p+}	**0.22029**	0.04185	**0.33863**	0.08313	**0.32689**	0.07471	**0.39025**	0.18467	**0.28517**	0.13060
β_{p+}^{p-}	**0.09031**	0.01303	**0.08554**	0.01513	**0.10803**	0.02455	**0.10739**	0.03644	**0.11533**	0.04229
β_{p-}^{p-}	**0.12089**	0.00990	**0.07924**	0.00846	**0.10560**	0.01492	**0.11708**	0.01852	**0.12703**	0.01769
β_{v}^{p-}	**0.18778**	0.03254	**0.24841**	0.06815	**0.14623**	0.05293	**1.99858**	0.54406	**0.71286**	0.24029
β_{p+}^{v}	**0.22012**	0.05897	**0.58323**	0.13571	**0.62347**	0.16982	**0.82809**	0.17549	**0.46870**	0.11512
β_{p-}^{v}	**0.52030**	0.10805	**0.40219**	0.19699	**1.15156**	0.22189	**1.64153**	0.27352	**0.50852**	0.20679
β_{v}^{v}	**0.75158**	0.34721	**0.50411**	0.12457	**0.56593**	0.27920	**0.89291**	0.41545	**0.91520**	0.43494

Table 2.16: Posterior Mean and Standard Deviation of Parameters in a 3-dimensional Hawkes Structure Model (without impact of marks and before de-periodisation)

	FCHI		IBEX		SPDR	
	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.
λ_0^{p+}	**0.004**	0.000	**0.003**	0.000	**0.002**	0.000
λ_0^{p-}	**0.003**	0.000	**0.003**	0.000	**0.002**	0.000
λ_0^{v}	**0.004**	0.000	**0.004**	0.000	**0.002**	0.000
ϑ_{p+}^{p+}	0.002	0.002	0.006	0.003	**0.078**	0.009
ϑ_{p-}^{p+}	0.002	0.002	**0.045**	0.008	**0.010**	0.004
ϑ_{v}^{p+}	0.003	0.002	0.003	0.002	0.008	0.006
ϑ_{p+}^{p-}	0.001	0.001	**0.034**	0.006	0.006	0.004
ϑ_{p-}^{p-}	**0.021**	0.006	**0.038**	0.008	**0.082**	0.014
ϑ_{v}^{p-}	0.002	0.001	0.004	0.003	0.008	0.005
ϑ_{p+}^{v}	0.001	0.001	**0.018**	0.006	0.004	0.004
ϑ_{p-}^{v}	0.002	0.002	**0.018**	0.008	**0.014**	0.007
ϑ_{v}^{v}	0.001	0.001	**0.016**	0.007	**0.025**	0.009
β_{p+}^{p+}	**0.747**	0.185	**0.888**	0.185	**0.863**	0.062
β_{p-}^{p+}	**1.304**	0.110	**1.112**	0.155	**0.899**	0.144
β_{v}^{p+}	**0.725**	0.197	**0.732**	0.208	**0.703**	0.199
β_{p+}^{p-}	**0.777**	0.141	**1.062**	0.117	**0.664**	0.129
β_{p-}^{p-}	**0.567**	0.100	**0.548**	0.075	**0.793**	0.080
β_{v}^{p-}	**1.318**	0.373	**1.389**	0.360	**1.262**	0.328
β_{p+}^{v}	**0.702**	0.204	**1.148**	0.167	**0.653**	0.187
β_{p-}^{v}	**0.607**	0.177	**0.927**	0.155	**0.714**	0.154
β_{v}^{v}	**0.813**	0.234	**1.290**	0.172	**1.067**	0.175

Notes: The posterior mean and standard deviation of parameters presented here are from Hawkes Process kernel without impact of marks and before the de-periodisation of asset returns, λ_0 denotes ground intensity, ϑ_q^i denotes average number jumps in dimension i directly induced by jumps in dimension q and β_q^i denotes corresponding decaying speeds. Values that are significant greater than 0 are marked bold.

Table 2.17: Posterior Mean and Standard Deviation of Parameters in a 3-dimensional Hawkes Structure Model (without impact of marks and after de-periodisation)

	FCHI		IBEX		SPDR	
	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.
λ_0^{p+}	**0.003**	0.000	**0.003**	0.000	**0.001**	0.000
λ_0^{p-}	**0.004**	0.000	**0.004**	0.000	**0.001**	0.000
λ_0^{v}	**0.000**	0.000	**0.001**	0.000	**0.000**	0.000
ϑ_{p+}^{p+}	**0.112**	0.013	**0.091**	0.012	**0.100**	0.021
ϑ_{p-}^{p+}	**0.093**	0.015	**0.059**	0.010	**0.043**	0.015
ϑ_{v}^{p+}	**0.149**	0.025	**0.189**	0.022	**0.126**	0.027
ϑ_{p+}^{p-}	**0.042**	0.010	**0.050**	0.012	**0.065**	0.018
ϑ_{p-}^{p-}	**0.211**	0.016	**0.197**	0.017	**0.165**	0.024
ϑ_{v}^{p-}	**0.164**	0.027	**0.119**	0.021	0.026	0.015
ϑ_{p+}^{v}	**0.007**	0.003	**0.015**	0.005	**0.015**	0.007
ϑ_{p-}^{v}	**0.019**	0.005	**0.016**	0.005	**0.035**	0.012
ϑ_{v}^{v}	0.026	0.016	**0.070**	0.021	**0.073**	0.034
β_{p+}^{p+}	**0.219**	0.025	**0.219**	0.028	**0.209**	0.039
β_{p-}^{p+}	**0.124**	0.026	**0.222**	0.039	**0.227**	0.081
β_{v}^{p+}	**0.397**	0.074	**0.548**	0.077	**0.420**	0.087
β_{p+}^{p-}	**0.185**	0.048	**0.152**	0.033	**0.165**	0.039
β_{p-}^{p-}	**0.165**	0.012	**0.165**	0.013	**0.188**	0.023
β_{v}^{p-}	**0.482**	0.102	**0.526**	0.112	**1.051**	0.327
β_{p+}^{v}	**0.532**	0.106	**0.240**	0.055	**0.646**	0.137
β_{p-}^{v}	**0.247**	0.051	**0.410**	0.114	**0.325**	0.084
β_{v}^{v}	**0.581**	0.205	**0.564**	0.119	**0.788**	0.184

Notes: The posterior mean and standard deviation of parameters presented here are from Hawkes Process kernel without impact of marks and before the de-periodisation of asset returns, λ_0 denotes ground intensity, ϑ_q^i denotes average number jumps in dimension i directly induced by jumps in dimension q and β_q^i denotes corresponding decaying speeds. Values that are significant greater than 0 are marked bold.

Table 2.18: Posterior of Parameters of Marked Hawkes kernel in $\mathcal{M}_4$ with Impact Function in (3,I)

	FCHI		IBEX		SPDR	
	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.
λ_0^{p+}	**0.003**	0.000	**0.003**	0.000	**0.001**	0.000
λ_0^{p-}	**0.004**	0.000	**0.004**	0.000	**0.001**	0.000
λ_0^{v}	**0.001**	0.000	**0.001**	0.000	**0.000**	0.000
ϑ_{p+}^{p+}	**0.112**	0.018	**0.091**	0.017	**0.101**	0.031
ϑ_{p-}^{p+}	**0.092**	0.022	**0.058**	0.018	**0.046**	0.016
ϑ_{v}^{p+}	**0.056**	0.040	**0.060**	0.034	**0.049**	0.042
ϑ_{p+}^{p-}	**0.035**	0.017	**0.042**	0.017	**0.062**	0.026
ϑ_{p-}^{p-}	**0.217**	0.024	**0.200**	0.023	**0.168**	0.038
ϑ_{v}^{p-}	**0.123**	0.043	**0.055**	0.044	**0.013**	0.010
ϑ_{p+}^{v}	0.006	0.004	0.012	0.008	0.015	0.009
ϑ_{p-}^{v}	0.019	0.007	0.016	0.008	0.037	0.020
ϑ_{v}^{v}	**0.011**	0.009	**0.019**	0.014	0.028	0.018
$\tilde{\alpha}_{p+}^{p+}$	**0.959**	0.122	**0.963**	0.152	**0.932**	0.253
$\tilde{\alpha}_{p-}^{p+}$	**0.940**	0.178	**0.733**	0.195	**0.288**	0.202
$\tilde{\alpha}_{v}^{p+}$	**0.924**	0.678	**0.927**	0.545	**0.963**	0.833
$\tilde{\alpha}_{p+}^{p-}$	**0.788**	0.300	**0.817**	0.268	**0.829**	0.313
$\tilde{\alpha}_{p-}^{p-}$	**0.939**	0.084	**0.940**	0.092	**0.818**	0.156
$\tilde{\alpha}_{v}^{p-}$	**0.986**	0.341	**0.935**	0.755	**0.949**	0.737
$\tilde{\alpha}_{p+}^{v}$	**0.722**	0.415	0.779	0.408	0.781	0.473
$\tilde{\alpha}_{p-}^{v}$	**0.820**	0.264	0.451	0.290	0.731	0.318
$\tilde{\alpha}_{v}^{v}$	**0.959**	0.769	**0.934**	0.691	0.973	0.614
$\tilde{\beta}_{p+}^{p+}$	19.991	17.657	**17.786**	16.553	**27.412**	20.455
$\tilde{\beta}_{p-}^{p+}$	**28.016**	27.435	**125.783**	56.375	**280.218**	61.124
$\tilde{\beta}_{v}^{p+}$	**66.099**	30.008	**60.078**	19.427	**26.926**	18.916
$\tilde{\beta}_{p+}^{p-}$	104.166	84.914	89.332	68.702	69.072	47.433
$\tilde{\beta}_{p-}^{p-}$	**28.103**	11.602	**28.178**	10.695	**71.506**	28.584
$\tilde{\beta}_{v}^{p-}$	**12.041**	8.697	**53.781**	32.445	**37.745**	24.673
$\tilde{\beta}_{p+}^{v}$	134.740	113.922	107.709	115.620	89.732	61.755
$\tilde{\beta}_{p-}^{v}$	81.606	55.260	260.896	90.044	106.356	92.986
$\tilde{\beta}_{v}^{v}$	35.231	26.852	54.180	26.875	**19.788**	16.964
$\tilde{\gamma}_{p+}^{p+}$	**167.724**	65.677	234.029	139.398	**289.586**	108.182
$\tilde{\gamma}_{p-}^{p+}$	141.825	118.879	925.977	141.438	877.951	338.995
$\tilde{\gamma}_{v}^{p+}$	140.530	70.431	126.067	41.242	221.596	55.311
$\tilde{\gamma}_{p+}^{p-}$	140.492	135.005	563.262	174.851	585.436	226.470
$\tilde{\gamma}_{p-}^{p-}$	258.639	31.179	161.399	107.240	264.687	46.598
$\tilde{\gamma}_{v}^{p-}$	**176.232**	40.447	**113.148**	70.907	**147.470**	83.924
$\tilde{\gamma}_{p+}^{v}$	989.742	454.925	777.175	539.579	158.086	122.841
$\tilde{\gamma}_{p-}^{v}$	**1336.195**	242.598	**836.646**	426.098	**140.314**	109.936
$\tilde{\gamma}_{v}^{v}$	**591.180**	332.004	**178.997**	122.933	**241.731**	99.832
β_{p+}^{p+}	**0.222**	0.027	**0.223**	0.031	**0.223**	0.052
β_{p-}^{p+}	**0.128**	0.028	**0.240**	0.046	**0.272**	0.059
β_{v}^{p+}	**0.470**	0.115	**0.604**	0.107	**0.518**	0.150
β_{p+}^{p-}	**0.262**	0.092	**0.200**	0.073	**0.185**	0.054
β_{p-}^{p-}	**0.165**	0.013	**0.164**	0.015	**0.189**	0.023
β_{v}^{p-}	**0.394**	0.077	**0.357**	0.076	**1.198**	0.353
β_{p+}^{v}	**0.700**	0.167	**0.353**	0.154	**0.571**	0.205
β_{p-}^{v}	**0.290**	0.054	**0.462**	0.103	**0.375**	0.077
β_{v}^{v}	**0.587**	0.204	**0.752**	0.119	**0.750**	0.168

Table 2.19: Posterior Mean and Standard Deviation of Parameters of Marked Hawkes kernel in $\mathcal{M}_5$ with Impact Function in (3,II)

	FCHI		IBEX		SPDR	
	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.
λ_0^{p+}	**0.003**	0.000	**0.003**	0.000	**0.001**	0.000
λ_0^{p-}	**0.004**	0.000	**0.004**	0.000	**0.001**	0.000
λ_0^{v}	**0.000**	0.000	**0.001**	0.000	**0.000**	0.000
ϑ_{p+}^{p+}	**0.083**	0.011	**0.078**	0.012	**0.087**	0.022
ϑ_{p-}^{p+}	**0.078**	0.012	**0.051**	0.009	**0.034**	0.012
ϑ_{v}^{p+}	**85.452**	30.895	**192.658**	61.408	**142.890**	70.663
ϑ_{p+}^{p-}	**0.031**	0.009	**0.041**	0.010	**0.053**	0.016
ϑ_{p-}^{p-}	**0.158**	0.013	**0.159**	0.016	**0.122**	0.020
ϑ_{v}^{p-}	**138.764**	46.791	**94.184**	33.300	**38.076**	23.512
ϑ_{p+}^{v}	**0.004**	0.002	**0.011**	0.004	0.014	0.005
ϑ_{p-}^{v}	**0.016**	0.004	**0.011**	0.003	**0.036**	0.009
ϑ_{v}^{v}	**0.674**	0.668	**19.850**	9.338	**0.000**	0.000
$\tilde{\alpha}_{p+}^{p+}$	**1.318**	0.031	**1.348**	0.039	**1.322**	0.064
$\tilde{\alpha}_{p-}^{p+}$	**1.544**	0.050	**1.482**	0.054	**1.778**	0.160
$\tilde{\alpha}_{v}^{p+}$	**-1.024**	-0.034	**-1.364**	-0.036	**-1.244**	-0.061
$\tilde{\alpha}_{p+}^{p-}$	**1.717**	0.121	**1.665**	0.098	**1.598**	0.112
$\tilde{\alpha}_{p-}^{p-}$	**1.207**	0.017	**1.196**	0.020	**1.275**	0.038
$\tilde{\alpha}_{v}^{p-}$	**-1.104**	-0.033	**-0.984**	-0.033	**-0.600**	-0.049
$\tilde{\alpha}_{p+}^{v}$	**2.068**	0.429	**1.893**	0.209	**1.595**	0.160
$\tilde{\alpha}_{p-}^{v}$	**1.837**	0.112	**1.686**	0.122	**1.441**	0.070
$\tilde{\alpha}_{v}^{v}$	**0.255**	0.135	**-0.629**	-0.033	**22.022**	-9.326
β_{p+}^{p+}	**0.202**	0.026	**0.195**	0.029	**0.214**	0.051
β_{p-}^{p+}	**0.111**	0.023	**0.202**	0.038	**0.143**	0.062
β_{v}^{p+}	**0.375**	0.071	**0.507**	0.082	**0.426**	0.128
β_{p+}^{p-}	**0.157**	0.067	**0.139**	0.035	**0.153**	0.052
β_{p-}^{p-}	**0.154**	0.013	**0.157**	0.016	**0.180**	0.027
β_{v}^{p-}	**0.303**	0.062	**0.289**	0.053	**0.190**	0.074
β_{p+}^{v}	**0.443**	0.142	**0.261**	0.093	**0.601**	0.174
β_{p-}^{v}	**0.227**	0.040	**0.496**	0.157	**0.371**	0.066
β_{v}^{v}	**0.626**	0.282	**0.417**	0.101	**0.217**	0.154

Table 2.20: Posterior Mean and Standard Deviation of Parameters of Marked Hawkes kernel in $\mathcal{M}_5$ with Impact Function in (3,III)

	FCHI		IBEX		SPDR	
	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.
λ_0^{p+}	**0.004**	0.000	**0.002**	0.000	**0.000**	0.000
λ_0^{p-}	**0.006**	0.000	**0.004**	0.000	**0.002**	0.000
λ_0^{v}	**0.001**	0.000	**0.001**	0.000	**0.000**	0.000
ϑ_{p+}^{p+}	**0.000**	0.000	**0.000**	0.000	**0.000**	0.000
ϑ_{p-}^{p+}	**0.000**	0.000	**0.000**	0.000	**0.000**	0.000
ϑ_{v}^{p+}	**0.000**	0.000	**0.000**	0.000	**0.000**	0.000
ϑ_{p+}^{p-}	**0.000**	0.000	**0.000**	0.000	**0.000**	0.000
ϑ_{p-}^{p-}	**0.000**	0.000	**0.000**	0.000	**0.000**	0.000
ϑ_{v}^{p-}	**0.000**	0.000	**0.000**	0.000	**0.000**	0.000
ϑ_{p+}^{v}	**0.000**	0.000	**0.000**	0.000	**0.000**	0.000
ϑ_{p-}^{v}	**0.000**	0.000	**0.000**	0.000	**0.000**	0.000
ϑ_{v}^{v}	**0.000**	0.000	**0.000**	0.000	**0.000**	0.000
$\tilde{\alpha}_{p+}^{p+}$	**31.511**	12.774	**67.592**	11.938	**15.903**	8.277
$\tilde{\alpha}_{p-}^{p+}$	**80.099**	20.888	**25.586**	17.731	**43.478**	36.040
$\tilde{\alpha}_{v}^{p+}$	98.249	78.788	**208.437**	46.617	180.038	5.861
$\tilde{\alpha}_{p+}^{p-}$	**46.316**	8.615	**59.871**	11.480	**19.686**	20.426
$\tilde{\alpha}_{p-}^{p-}$	**54.922**	14.958	**59.317**	8.152	**23.822**	14.449
$\tilde{\alpha}_{v}^{p-}$	23.828	15.754	71.632	68.347	50.547	23.737
$\tilde{\alpha}_{p+}^{v}$	66.493	16.519	**59.378**	20.031	**40.302**	14.454
$\tilde{\alpha}_{p-}^{v}$	**45.470**	14.235	**23.994**	9.607	89.200	14.957
$\tilde{\alpha}_{v}^{v}$	**24.577**	15.399	**128.280**	74.043	**169.166**	14.789
β_{p+}^{p+}	0.094	0.097	0.000	0.000	**0.000**	0.000
β_{p-}^{p+}	**0.024**	0.027	0.050	0.048	**1.835**	0.696
β_{v}^{p+}	0.116	0.120	0.255	0.094	0.252	0.045
β_{p+}^{p-}	0.407	0.229	0.141	0.105	0.434	0.123
β_{p-}^{p-}	0.032	0.031	0.000	0.000	**0.181**	0.235
β_{v}^{p-}	**0.456**	0.345	0.288	0.188	**1.293**	0.346
β_{p+}^{v}	0.198	0.136	0.402	0.153	**0.437**	0.250
β_{p-}^{v}	0.071	0.052	0.216	0.150	0.175	0.118
β_{v}^{v}	0.476	0.356	0.263	0.158	**0.434**	0.138

Table 2.21: Posterior Mean and Standard Deviation of Parameters of Marked Hawkes kernel in $\mathcal{M}_7$ with Impact Function in (3,IV)

	FCHI		IBEX		SPDR	
	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.
λ_0^{p+}	**0.003**	0.000	**0.003**	0.000	**0.001**	0.000
λ_0^{p-}	**0.004**	0.000	**0.004**	0.000	**0.001**	0.000
λ_0^{v}	**0.001**	0.000	**0.001**	0.000	**0.000**	0.000
ϑ_{p+}^{p+}	**0.112**	0.016	**0.091**	0.015	**0.102**	0.030
ϑ_{p-}^{p+}	**0.093**	0.021	**0.058**	0.024	**0.046**	0.021
ϑ_{v}^{p+}	**0.070**	0.047	**0.050**	0.035	**0.052**	0.034
ϑ_{p+}^{p-}	**0.037**	0.015	**0.044**	0.017	**0.060**	0.026
ϑ_{p-}^{p-}	**0.217**	0.022	**0.209**	0.028	**0.161**	0.032
ϑ_{v}^{p-}	**0.102**	0.058	**0.033**	0.025	**0.016**	0.013
ϑ_{p+}^{v}	0.006	0.004	0.014	0.008	0.015	0.008
ϑ_{p-}^{v}	0.020	0.010	0.016	0.008	0.037	0.015
ϑ_{v}^{v}	0.008	0.006	**0.030**	0.020	0.027	0.019
$\tilde{\alpha}_{p+}^{p+}$	**0.969**	0.121	**0.972**	0.145	**0.901**	0.233
$\tilde{\alpha}_{p-}^{p+}$	**0.939**	0.176	**0.762**	0.216	**0.432**	0.279
$\tilde{\alpha}_{v}^{p+}$	**0.946**	0.647	**0.890**	0.679	**0.964**	0.639
$\tilde{\alpha}_{p+}^{p-}$	**0.854**	0.294	**0.788**	0.269	**0.828**	0.284
$\tilde{\alpha}_{p-}^{p-}$	**0.945**	0.080	**0.897**	0.089	**0.788**	0.145
$\tilde{\alpha}_{v}^{p-}$	**0.968**	0.554	**0.964**	0.700	**0.946**	0.761
$\tilde{\alpha}_{p+}^{v}$	0.630	0.391	0.776	0.368	0.580	0.379
$\tilde{\alpha}_{p-}^{v}$	0.706	0.295	0.387	0.256	0.719	0.265
$\tilde{\alpha}_{v}^{v}$	0.949	0.729	**0.967**	0.635	0.963	0.664
$\tilde{\beta}_{p+}^{p+}$	15.029	12.261	**14.103**	11.749	**40.901**	25.159
$\tilde{\beta}_{p-}^{p+}$	**28.766**	22.435	**114.227**	89.751	**225.554**	75.350
$\tilde{\beta}_{v}^{p+}$	46.936	23.724	90.428	20.642	26.283	14.013
$\tilde{\beta}_{p+}^{p-}$	71.969	52.118	105.040	64.317	71.047	63.593
$\tilde{\beta}_{p-}^{p-}$	**25.674**	8.766	**49.204**	21.485	**84.030**	22.107
$\tilde{\beta}_{v}^{p-}$	27.901	14.912	29.798	33.365	39.667	31.935
$\tilde{\beta}_{p+}^{v}$	182.262	137.057	111.030	102.291	173.228	50.529
$\tilde{\beta}_{p-}^{v}$	137.855	89.911	293.846	108.216	111.663	57.667
$\tilde{\beta}_{v}^{v}$	**44.341**	30.037	**26.862**	18.316	**27.375**	20.229
β_{p+}^{p+}	**0.221**	0.025	**0.227**	0.032	**0.215**	0.044
β_{p-}^{p+}	**0.127**	0.028	**0.233**	0.048	**0.234**	0.057
β_{v}^{p+}	**0.428**	0.075	**0.592**	0.149	**0.571**	0.185
β_{p+}^{p-}	**0.237**	0.073	**0.187**	0.051	**0.198**	0.056
β_{p-}^{p-}	**0.164**	0.013	**0.164**	0.015	**0.191**	0.022
β_{v}^{p-}	**0.411**	0.119	**0.429**	0.126	**0.988**	0.182
β_{p+}^{v}	**0.926**	0.249	**0.297**	0.090	**0.896**	0.228
β_{p-}^{v}	**0.291**	0.061	**0.527**	0.104	**0.372**	0.074
β_{v}^{v}	**1.751**	0.337	**0.541**	0.178	**0.802**	0.189

Chapter 3

Jump Clustering and High-Frequency Forecasting on Volatility and Risk Measures

3.1 Introduction

Financial market jumps refer to significant discontinuities in the processes of price and variance. Its importance has been highlighted in studying stochastic volatility models for many years (see e.g., Duffie et al. 2000, Eraker 2004, Jacod & Todorov 2010, Todorov & Tauchen 2011). Early studies adopted an assumption that jumps arrive independently, while recent studies (e.g., Aït-Sahalia et al. 2015) reveal that jump activities have a dependent structure where jumps in one dimension[1] can raise the probability of future jumps in both its own dimension and other dimensions and therefore they model jumps by a Hawkes Process. While these models are based on daily frequency data, dynamics of price and variance on an intraday level has also attracted much attention in recent years due to the availability of high-frequency

[1]In this chapter, we specify three dimensions of jumps in the model - positive return jumps, negative return jumps and variance jumps. See details in the model specification section.

data (see e.g., Stroud & Johannes 2014, Chen et al. 2021).

However, how useful in practice are these different specifications of models? Does the dependent structure of jumps, or jump clustering feature, help forecast volatility and other risk measures? Are different jumps contributing to the forecast in different ways? This article aims to answer these questions, which are important to financial practitioners who require information on future market movements.

Hawkes Process (HP), proposed by Hawkes (1971*a*,*b*) as a self-exciting process, was initially used to model unexpected shocks and has attracted applications in finance recently. For example, Aït-Sahalia et al. (2015) incorporate a multivariate HP into an Itô semimartingale process and study cross-impact among international stock markets. Li & Zinna (2018) propose a pricing model for asset returns and variance swap rates, and find strong evidence of price and variance co-jumps and self-exciting behaviours of jumps. These papers study dynamics of daily data, while Boswijk et al. (2018) propose a model-free test on self-excitation in high-frequency jumps. Chen et al. (2021) look at 5-minute frequency data and propose a stochastic volatility model with self-exciting jumps. They find that self-exciting jumps that occur during the day can form clusters lasting between 3 to 20 trading hours. In addition, Stroud & Johannes (2014) propose a stochastic volatility model incorporating both price and volatility jumps, periodic effects and announcement effects.

In this chapter, we propose a forecasting framework under a stochastic volatility model featuring jump clustering and employ high-frequency intraday data. We use a Bayesian Markov chain Monte Carlo (MCMC) algorithm to estimate parameters in the in-sample period as is commonly adopted by literature (e.g. Eraker 2004, Stroud & Johannes 2014). In the out-of-sample period, a particle filter to forward simulate return and variance distribution. From simulated return and variance distributions, we obtain intraday point forecasts from 5 minutes to 6.5 hours (one trading day) for variance, realised volatility, Value-at-Risk (VaR) and expected shortfall (ES). To our knowledge, we are the first to consider jump clustering in intraday risk management and high-frequency forecasting.

In addition, we show estimated parameters in the stochastic volatility model are affected by microstructure noise in the high-frequency data. Forecasts by forward simulations are subjected to accumulated forecast errors due to the inconsistent estimated parameters, especially when forecasting with long forecasting horizons. To circumvent this issue, we implement the Kalman filter from forecasting in the meteorology field to correct the forecast errors in this chapter. We show the accumulation of forecast errors across forecast horizons due to the microstructure noise through a simulation study and show the usefulness of the Kalman filter in correcting this error. This is our another novel contribution.

Under this forecasting framework, we find strong evidence that the inclusion of jump clustering improves high-frequency volatility forecasting using an individual stock and a broad index data. We examine a range of forecasting horizons and a model with jump clustering features outperforms all other models across all forecast horizons in variance forecasting, and the forecast performance is comparatively better with forecast horizons less than 2 hours. We also find that the forecasting performance after arrivals of negative return jumps is especially better than the performance after arrivals of other types of jumps.

In daily realised volatility forecasting, we benchmark two popular classes of realised volatility forecasting models (HAR-class and Realised-GARCH class) and show our model outperforms benchmark models by comparing MSE, forecast bias, DM test results and through regression analysis. This confirms that intraday jump clustering features provide additional information in forecasts of lower frequency, i.e. daily realised volatility.

In VaR and ES forecasting, our model reports the lowest expected loss by using a loss function proposed by Fissler & Ziegel (2016); joint expected loss of VaR and ES can be reduced by up to 15% when including jump clustering in a model. We also adopt two popular backtesting methods and show VaR and ES violations of our model forecasts are least likely serially correlated. Noticeably, we also look at conditional performance by extracting forecasts after occurrences of different

jumps. Similar to variance forecasting, we find negative jump clustering adds more value than other types of jump do when using higher-frequency data and shorter forecasting horizons. We find that the benefits of including jump clustering in a model is higher with an individual stock data which contains more jumps than a broad index data, suggesting that our model works better when time-series with large numbers of jumps.

The rest of this chapter is organised as follows. Section 3.2 introduces our model. Section 3.3 presents in-sample estimation, out-of-sample forecasting and error correction approaches. Section 3.5 presents our empirical results. Section 3.6 concludes the chapter. Some technical results are confined to the appendix B.5.

3.2 Model Specifications

3.2.1 Return and Variance Process

We consider 5-minute high-frequency stock market data with $M = 78$ observations in a trading day (6.5 hours). Letting P_t denote the natural logarithm of an asset price at time t, we model the asset return at time t as $Y_t = P_t - P_{t-1}$, which follows:

$$Y_t \cdot s_t = \mu + \sqrt{V_{t-1}}\epsilon_t^Y + \xi_t^{Y+} J_t^{Y+} + \xi_t^{Y-} J_t^{Y-}, \tag{3.1}$$

where $s_t = I_{tk}\hat{f}_k^{WSD}$, $k = 1, ..., M$ is a periodic component, I_{tk} is a period indicator and $I_{tk} = 1$ when time t corresponds to period k. $\hat{f}_k^{WSD}$ is a weighted standard deviation estimator to adjust return for intraday periodicity (see Section 3.3.1). μ denotes the drift term of return, ϵ_t^Y is an identical independent distributed random variable following $\mathcal{N}(0, 1)$. We separate positive and negative return jump components to capture dynamics of their underlying intensities, $\{\xi_t^{Y+}, \xi_t^{Y-}\}$ denote return jump sizes with truncated normal distributions $\xi_t^{Y+} \sim \mathcal{N}(\mu_{Y+}, \sigma_{Y+})\mathbf{1}_{\xi^{Y+}>0}$, $\xi_t^{Y-} \sim \mathcal{N}(\mu_{Y-}, \sigma_{Y-})\mathbf{1}_{\xi^{Y-}<0}$ and $\{J_t^{Y+}, J_t^{Y-}\}$ are return jump indicators with underlying probabilities $P(J_t^i = 1) = \lambda_t^i$, which will be introduced in Section 3.2.2. V_t

denotes the variance process of the return, which evolve via

$$V_t = V_{t-1} + \kappa(\theta - V_{t-1}) + \sigma_V \sqrt{V_{t-1}} \epsilon_t^V + \xi_t^V J_t^V, \tag{3.2}$$

where κ and θ denote the mean reversion speed and long-run variance mean, respectively, and σ_V refers to the volatility of volatility, ϵ_t^V is a random variable in the process that follows $\mathcal{N}(0,1)$. We further let random variables in these two processes be correlated with a correlation $corr(\epsilon_t^Y, \epsilon_t^V) = \rho$. Similar to return jump components, ξ_t^V denotes the size of variance jump with an exponential distribution $\xi_t^V \sim exp(\mu_V)$, J_t^V denotes the variance jump indicator. At this stage, our model resembles the specifications of most jump-diffusion models (e.g. Duffie et al. (2000);Eraker (2004);Stroud & Johannes (2014);Aït-Sahalia et al. (2015)).

3.2.2 Marked Hawkes Process

Early literature adopts a simple assumption on jump components, that jumps are serially independent, which makes the underlying intensity a constant value ($\lambda_t^i \equiv \lambda_0$). In this chapter, we assume a multivariate marked Hawkes process (MHP) on λ_t^i to capture the dynamic dependent structure among jump components. λ_t^i is specified as follows:

$$\lambda_t^i = \lambda_0^i + \sum_q \vartheta_{q,i} \sum_{0<\tilde{t}<t} \phi_{q,i}(t-\tilde{t}) \omega_{q,i}(\xi^q), \quad i,q \in \{Y+, Y-, V\}, \tag{3.3}$$

where λ_0^i denotes a base line intensity of dimension i, which is constant. $\boldsymbol{\vartheta}$ is a branching coefficient matrix. It governs the mean increase of intensity to the process in dimension i that is produced by the events in dimension q. $\phi_{q,i}$ is a decay function such that $\phi : \mathbb{R}_+ \to \mathbb{R}_+$, and we assume an exponential decaying kernel $\phi_{q,i}(t-\tilde{t}) = \beta_{q,i} e^{-\beta_{q,i}(t-\tilde{t})}, \beta_{q,i} > 0$. $\omega_{q,i}(\xi^q)$ is a non-decreasing impact function of jump sizes, such that $\omega : \mathbb{R} \to \mathbb{R}_+$. We adopt a normalised impact function as follows:

$$\omega_{q,i}(\xi^q) = \frac{\tilde{\omega}_{q,i}(\xi^q)}{\mathbb{E}[\tilde{\omega}_{q,i}(\xi^q)]}, \quad \tilde{\omega}_{q,i}(\xi^q) = \tilde{\alpha}_{q,i} + \tilde{\beta}_{q,i} log(1 + |\xi^q|) \tag{3.4}$$

For a simple example, the underlying intensity of dimension i at time t (λ_t^i) equals a constant value λ_0^i when there are no jumps. If a jump occurs in dimension q with size ξ^q, responding to that jump, the intensity in dimension i will increase by $\vartheta_{q,i} \cdot \omega_{q,i}(\xi^q)$, where $\omega_{q,i}(\xi^q)$ captures the impact of different sizes of jump sizes. This incremental intensity will decay at a speed of $\phi_{q,i}(1) = \beta_{q,i} e^{-\beta_{q,i} \cdot 1}$ every time interval. See a detailed interpretation and explanation of MHP by Liniger (2009).

3.3 Model Estimation and Forecasting

In this section, we discuss our non-parametric estimations[2] on intraday periodic component s_t and jump detection J_t. Then we discuss our Bayesian inference on parameters and latent variables in the in-sample period. In the out-of-sample period, we adopt a particle filter to estimate latent state and forward simulate distributions of return and volatility. We also use a Kalman filter to adjust simulated distributions.

3.3.1 Intraday Periodicity

The theory of intraday periodicity in high-frequency data has been well-established (see, e.g., Andersen & Bollerslev (1997)). Chen et al. (2021) show that intraday periodicity, especially overnight returns, can strongly hinder jump clustering features. Boudt et al. (2011) show that intraday periodicity can improve the accuracy of jump detection. We follow them and apply a weighted standard deviation estimator in

[2]The reason for conducting non-parametric estimations on s_t and J_t is to improve the efficiency in the Bayesian estimation. Due to the inclusion of MHP parameters, we found difficulties in parameters convergence due to a large number of latent variables being estimated simultaneously, while these non-parametric estimations partially solve the problem. Additionally, we conduct simulations on this issue and try both approaches in simulated data. The power of jump detection under our non-parametric method is even higher.

the model. We firstly transform $\{Y_t\}$ to an $(M \times T)$ matrix, and define the order statistics of returns $\bar{Y}_{(1),k} \leq \bar{Y}_{(2),k} \leq ... \leq \bar{Y}_{(T),k}$ for $k = 1, ..., M$. We further define a shortest half-scale statistic as follows:

$$ShortH_k = 0.741 \cdot min\{\bar{Y}_{(h'),k} - \bar{Y}_{(1),k}, ..., \bar{Y}_{(T),k} - \bar{Y}_{(T-h'+1),k}\}, \tag{3.5}$$

where $h' = \frac{T}{2}+1$, this statistics is essentially the minimum differences among return's halves. Then a shortest half-scale estimator is defined by:

$$\hat{f}_k^{ShortH} = \frac{ShortH_k}{\frac{1}{M}\sum_{k=1}^{M} ShortH_k^2}, \quad k = 1, ..., M, \tag{3.6}$$

and the weighted standard deviation estimator can be obtained as follows:

$$\hat{f}_k^{WSD} = \frac{WSD_k}{\frac{1}{M}\sum_{k=1}^{M} WSD_k^2}, \quad WSD_i = \sqrt{1.081 \cdot \frac{\sum_{k=1}^{T} w_{i,k}\bar{Y}_{i,k}^2}{\sum_{k=1}^{T} w_{i,k}}}, \quad i = 1, ..., T, \tag{3.7}$$

where $w_{i,k} = I(\bar{Y}_{i,k}/\hat{f}_k^{ShortH})$ is a weight function with a identification function $I(\cdot)$, such that $I(x) = 1$ if $x \leq 6.635$ and 0 otherwise.

3.3.2 Jump Detection

Denoting the return of an asset at time t adjusted by periodicity as $r_t = Y_t \cdot s_t$, our goal is estimating $\{J_t^Y, J_t^V\}$. Notice we further separate return jumps to positive and negative, but we only consider positive variance jumps since results show very few negative variance jumps. Additionally, downside movements of variance are captured by its mean-reversion structure (κ, θ).

There are plenty of techniques for detecting return jump activities in a finite time interval, and most detection methods rely on a continuous return variation estimator (see a comprehensive review by Maneesoonthorn et al. (2020)). We follow Mancini et al. (2015) and Figueroa-López & Mancini (2019) to construct our test. We identify a jump at time t, $J_t = 1$, when the squared return is greater than a threshold, $r_t^2 > \widehat{V_t^2} \cdot 2\Delta log\frac{1}{\Delta}$. $\widehat{V_t^2}$ is a non-parametric estimator of spot variance

based on pre-truncated returns:

$$\widehat{V_t^2} = \sum_{\tilde{t}=1}^{n} f_h(t-\tilde{t}) r_t^2 \cdot \mathbb{1}_{\{r_t^2 \leq 9\Delta^{0.99}\}} \tag{3.8}$$

where $f_h(\cdot)$ is weight function, $f_h(t) = \frac{1}{h} \cdot \frac{e^{-|t/h|}}{2}$ with a bandwidth $h = 200\Delta$ for simplicity. The idea of this filtering is to extract those standardised squared returns $(r_t\Delta/\widehat{V_t^2})$ which is not generated by a Brownian motion, whose absolute value is greater than the threshold $\sqrt{2log(1/\Delta)}$.

In filtering variance jumps, we follow Jacod & Todorov (2010) and set up test statistic for variance jumps as follow:

$$\mathcal{L}^v(t) = 2log(\frac{1}{2}(\widehat{V_t^2} + \widehat{V_{t-}^2})) - log(\widehat{V_t^2}) - log(\widehat{V_{t-}^2}) \tag{3.9}$$

and $N^b\mathcal{L}^v(t) \to \mathcal{X}_1^2$, where $b = \frac{1}{2} - \Delta t$. The $\widehat{V_t^2}$ is a spot variance estimator, and The temporal variance estimator $\widehat{V_t^2}$ is specified in Eq. 3.8.

3.3.3 Bayesian Inference on Parameters

In the in-sample period, we conduct a Bayesian inference on static parameter vector $\Theta = \{\mu, \kappa, \theta, \sigma_V, \rho, \mu_{Y+}, \sigma_{Y+}, \mu_{Y-}, \sigma_{Y-}, \mu_V, \lambda_0^i, \beta_{q,i}, \tilde{\alpha}_{q,i}, \tilde{\beta}_{q,i}\}$ for $i, q \in \{Y^+, Y^-, V\}$ and latent variables $\mathcal{Z}_t = \{V_t, \xi_t^{Y+}, \xi_t^{Y-}, \xi_t^V, \boldsymbol{\vartheta}\}$. Therefore, the joint posterior distribution is given by:

$$p(\Theta, \mathcal{Z}_t|Y_t) \propto p(Y_t|\mathcal{Z}_t)\, p(\mathcal{Z}_t|\Theta)\, p(\Theta) \tag{3.10}$$

We adopt the Markov chain Monte Carlo (MCMC) approach to simulate the posterior, the algorithm is a hybrid of Metropolis-Hastings and Gibbs sampling, since not all posterior distributions are available in closed forms. In terms of sampling parameters in MMHP and the branching matrix $\boldsymbol{\vartheta}$, we follow Rasmussen (2013) and construct our sampling steps. A detailed specification of priors and the algorithm is provided in Appendix B.1.

3.3.4 Out-of-Sample Forecasting Method

In the out-of-sample period, following Stroud & Johannes (2014), we firstly fix static parameters Θ at their posterior mean, which is $\widehat{\Theta}$ estimated in the in-sample period. Then, for every time t in out-of-sample, we firstly adopt a particle filter proposed by Pitt & Shephard (1999) to filter $\mathcal{Z}_t$ by sampling from $p\left(\mathcal{Z}_t|Y_t,\widehat{\Theta}\right)$. Then, we do out-of-sample forecasting by sampling from $p\left(\mathcal{Z}_{t+1}|\mathcal{Z}_t,\widehat{\Theta}\right)$ and $p\left(Y_{t+1}|\mathcal{Z}_{t+1},\widehat{\Theta}\right)$. Full details of particle filter sampling and re-sampling scheme is provided in Appendix B.2.

Specifically, we approximate forecast distributions of return and variance by forward simulations. Given a forecast horizon of τ [3] and the number of simulations of n, for $n_i = 1, ..., n$ and $s = 1, ..., \tau$, we firstly calculate underlying intensity λ^i_{t+s} using Equation 3.3, notice λ^i_{t+s} is calculated by summing up the impact of past jumps ($J^{i,(n_i)}_{[0,t+s]}$), for speeding up calculations, we only consider jumps in the past 20 trading days ($J^{i,(n_i)}_{[t+s-780,t+s]}$). Then jumps are sampled from a Bernoulli distribution:

$$J^{i,(n_i)}_{t+s} \sim Bernoulli(\lambda^i_{t+s}), \quad i = \{Y+, Y-, V\}, \tag{3.11}$$

and corresponding jump sizes ($\xi^{i,(n_i)}_{t+s}$) are sampled from their truncated normal and exponential distributions as mentioned in Section 3.2.1. Further, we simulate:

$$V^{(n_i)}_{t+s} \sim p\left(V^{(n_i)}_{t+s}|V^{(n_i)}_{t+s-1}, J^{V,(n_i)}_{t+s}, \xi^{V,(n_i)}_{t+s}, \widehat{\Theta}, x_{t+s}\right) \tag{3.12}$$

$$Y^{(n_i)}_{t+s} \sim p\left(Y^{(n_i)}_{t+s}|\mathcal{Z}^{(n_i)}_{t+s}, \widehat{\Theta}\right), \tag{3.13}$$

where x_{t+s} is an error correction term introduced in Section 3.4. We aggregate simulated variance and returns over forecast horizon τ. Note we also consider the

[3] Since we do forecasts of different frequencies by aggregating τ steps forward-simulated forecasts, 'τ forecast horizon' in this chapter also means one-step forecast under τ frequency. For a specific example, for 5-minute data, there are 78 observations in a trading day (6.5 hours). When we set $\tau = 6$ (0.5 hours), at time t, we aggregate forecasts at $t+1, t+2, ..., t+6$ to get forecast of $t+\tau$ at time t. Therefore, τ denotes the data or forecast frequency (0.5 hours), which is also a forecast horizon (6 small steps forecasts).

daily realised volatility measure proposed by Andersen & Bollerslev (1997), which is calculated by summing up intraday squared returns (setting $\tau = M$), so we simulate the distribution of realised volatility by:

$$V_{t,\tau}^{(n_i)} = \sum_{s=1}^{\tau} V_{t+s}^{(n_i)}, \quad Y_{t,\tau}^{(n_i)} = \sum_{s=1}^{\tau} Y_{t+s}^{(n_i)}, \quad \widehat{RV}_{t,\tau}^{(n_i)} = \sqrt{\sum_{s=1}^{\tau} (Y_{t+s}^{(n_i)})^2}, \tag{3.14}$$

[4] and point forecasts of variance and realised volatility is calculated by their posterior mean:

$$\widehat{V}_{t,\tau} = \frac{1}{n} \sum_{n_i=1}^{n} V_{t,\tau}^{(n_i)}, \quad \widehat{RV}_{t,\tau} = \frac{1}{n} \sum_{n_i=1}^{n} \widehat{RV}_{t,\tau}^{(n_i)} \tag{3.15}$$

Further, we consider two prevailing risk measures in the financial market, Value-at-Risk (VaR) and Expected Shortfall (ES), which are defined as follows:

$$VaR_t^{\alpha} \equiv inf\{Y_t \in \mathbb{R} | F_Y(Y_t|\mathcal{F}_{t-1}) \geq \alpha\}, \quad ES_t^{\alpha} \equiv \mathbb{E}[Y_t | Y_t \leq VaR_t^{\alpha}, \mathcal{F}_{t-1}]. \tag{3.16}$$

where $F_Y(\cdot|\mathcal{F}_{t-1})$ is a cumulative distribution of asset returns Y_t over a time horizon and under a significance level $\alpha \in (0, 1)$. In our forecasting approach, given the availability of posterior return distribution over horizon τ, we take α percentage quantile and posterior mean of simulations lower than α percentage quantile as our VaR and ES forecasts:

$$\widehat{VaR}_{t,\tau} = Y_{t,\tau}^{(o=\alpha*n)} \quad \widehat{ES}_{t,\tau} = \sum_{o=1}^{\alpha*n} Y_{t,\tau}^{(o)} \tag{3.17}$$

where o is an order indicator of n particles from $Y_{t,\tau}^{(n_i)}$. They are sorted in an ascending manner as $Y_{t,\tau}^{(o=1)}, ..., Y_{t,\tau}^{(o=n)}$. This is our novel contribution - including jump clustering in variance and risk forecasting, and examining its benefits under a high-frequency setting.

[4] $X_{t,\tau}$ denotes variable $\{X_t\}$'s aggregated value from time t to $t + \tau$, it is also $X_{t:t+\tau}$, this notation is used throughout this chapter.

3.4 Error Correction by Kalman Filter

The reason for correcting the simulated posterior distribution of variance by an error correction term is due to the potential accumulation of forecast errors (see Figure 3.1 plotting variance forecasts against a proxy actual values, the gap between them widens with increasing forecast horizons). We suspect these errors possibly come from microstructure noise (MN). More importantly, these errors can accumulate when aggregating forecasts over τ forecast horizons. We conduct a simulation study on resources of forecast errors and how they accumulate across forecast horizons in Section 3.4.3. However, we also highlight that how the nature of high-frequency data affects parameter estimations of stochastic volatility models is a big topic. In this chapter, we focus more on correcting these errors. To our knowledge, we are the first to do so.

Figure 3.1 plots SPX variance forecasts by the SVIJ-MHP model without error correction and its actual values under different forecast horizons. Clearly, the gap between forecasts and actual values widens with the increase of the forecast horizon. One could argue that one option to avoid this is to estimate the model with forecast-horizon-corresponded data frequency (e.g. estimate the model using daily data and forecast daily variance). However, this may lose intraday dynamics, especially intraday jump clustering features. Thus, we estimate the model with higher-frequency data and conduct a further error-correction step. Our error correction strategy assumes a continuous and auto-correlated component in the forecast errors. We aim to estimate and forecast this component iteratively. For example, at time $t + \tau$, we firstly estimate the continuous component in forecast errors (denoted as $x_{t,\tau}$ and $u_{t,\tau}$ respectively). Then, we forecast the continuous component for the next τ ($x_{t+\tau,\tau}$) from $t + \tau$ to $t + 2\tau$. Lastly, we compute x_t for each time point in this period $x_{t+\tau,1}, x_{t+\tau+1,1}, ..., x_{t+2\tau-1,1}$ to adjust forecasts of each time point. In this way, we can do variance forecasting with higher frequency data (e.g. forecast daily variance with 5-minute data) without losing the advantages of considering jump clustering

in the higher frequency.

[INSERT FIGURE 3.1 ABOUT HERE]

3.4.1 Kalman Filter

We are inspired by earlier papers in the field of meteorology (e.g. Evensen (1994)) which use Kalman filter (KF) to capture and correct forecast biases. KF regards forecast biases as a latent state in the error series. It assumes a serial linear correlation of biases and iteratively estimate and predict biases. We find it valuable in the financial forecasting field given a serial-correlated component in forecast errors. To our knowledge, we are the first to do so.

Denoting forecast error from t to $t+\tau$ as $u_{t,\tau}$ and the continuous term as $x_{t,\tau}$, our approach for correction is firstly letting x_t accumulate for τ forecasting horizon ($x_{t,\tau}$), then we use a KF to recursively update and predict $x_{t,\tau}, x_{t+\tau,\tau}, x_{t+2\tau,\tau}, ...$ Our goal is to forecast x_t for the next τ, $\widehat{x}_{t+\tau,\tau}$, given information at time $t+\tau$. We firstly let:

$$u_{t,\tau} = F \cdot x_{t,\tau} + \eta_{t,\tau}, \quad \eta_{t,\tau} \sim \mathcal{N}(0, \sigma_\eta^2) \tag{3.18}$$

we further let $x_{t,\tau}$ to be a linear function of its lag:

$$x_{t,\tau} = Q \cdot x_{t-\tau,\tau} + \delta_{t,\tau}, \quad \delta_{t,\tau} \sim \mathcal{N}(0, \sigma_\delta^2) \tag{3.19}$$

The initial state, $x_{0,\tau}$, is assumed to follow a normal distribution $\mathcal{N}(\mu_{x_{0,\tau}}, \sigma^2_{x_{0,\tau}})$. Then, the system of forecast errors forms a KF. In the update stage, the estimated continuous term $\widehat{x}_{t,\tau}$ is recursively updated by:

$$\widehat{x}_{t,\tau} = Q \cdot \widehat{x}_{t-\tau,\tau} + K_{t,\tau}(u_{t,\tau} - F \cdot \widehat{x}_{t-\tau,\tau}), \tag{3.20}$$

which is a linear combination of previously predicted $\widehat{x}_{t-\tau,\tau}$ and $u_{t,\tau}$. K_t is a weighting coefficient, called Kalman gain, which is given by:

$$K_{t,\tau} = \frac{F \cdot p_{t-\tau,\tau}}{F \cdot p_{t-\tau,\tau} + \sigma_\eta^2} \tag{3.21}$$

where $p_{t,\tau}$ is updated state error covariance, which encrypts the state error covariance that KF thinks the estimated error has. It is updated by:

$$p_{t,\tau} = p_{t-\tau,\tau}(1 - K_{t,\tau}) \tag{3.22}$$

Then, in the predictive stage, the forecast of x_t for the next τ is given by:

$$\widehat{x}_{t+\tau,\tau} = Q \cdot \widehat{x}_{t,\tau} \tag{3.23}$$

and state error covariance forecast is given by:

$$p_{t+\tau,\tau} = Q \cdot p_{t,\tau} + \sigma_\delta^2 \tag{3.24}$$

Notice this error correction requires the estimation of the initial state and parameters in the Kalman filter $\{\mu_{x_t}, \sigma_{x_t}, F, Q, \sigma_\eta, \sigma_\delta\}$. Therefore, we further separate the out-of-sample period to a validation period and the remaining out-of-sample period. In the validation period, we estimate the initial state and parameters of the Kalman filter using the expectation–maximisation (EM) algorithm proposed by Shumway & Stoffer (1982). In the remaining out-of-sample period, we do the out-of-sample forecasting.

3.4.2 Error Correction in Variance Forecasting with Longer Forecasting Horizons

This section introduces how we nest the KF in variance forecasting with a particle filter (Eq. 3.12).

If not considering the error correction, at time t, for forecast horizon $s = 1, ..., \tau$, we forward simulate variance by sampling from the predictive distribution:

$$V_{t+s}^{(n_i)} \sim P(V_{t+s}^{(n_i)} | V_{t+s-1}^{(n_i)}, J_{t+s}^{V,(n_i)}, \xi_{t+s}^{V,(n_i)}, \widehat{\Theta})$$

then, we aggregate forward simulated variance over τ for each particle and take particle means as point forecast ($\widehat{V}_{t,\tau}$) (Eq. 3.14, 3.15). However, We showed this

forecast is subjected to certain accumulated errors. Assuming the accumulated error term on $\widehat{V}_{t,\tau}$ is $x_{t,\tau}$. Section 3.4.1 introduces how we can estimate $\widehat{x}_{t-\tau,\tau}$ and forecast $\widehat{x}_{t,\tau}$ by KF. The question left is how we translate the estimated continuous errors over forecast horizon τ ($\widehat{x}_{t,\tau}$) to each step of variance forecast ($\widehat{V}_{t,s}$, $s = 1, 2..., \tau$). Therefore, we introduce the following proposition:

Proposition 1 *Given a variance process in Eq. 3.2 and a forecast horizon τ, if the forecast at time t, $\widehat{V}_{t,\tau}$ is biased by $\widehat{x}_{t,\tau}$, assuming forecast errors are accumulated evenly across the forecast horizon τ, then for each step across the forecast horizon $s = 1, 2, ..., \tau$, variance forecasts should be adjusted by $f(\tau) \cdot \widehat{x}_{t,\tau}$, where:*

$$f(\tau) = \kappa \left[1 + \tau - \frac{1 - (1 - \kappa)^{\tau+1}}{\kappa} \right]^{-1} \tag{3.25}$$

where κ denotes the mean-reversion speed parameter of the variance process (see Eq. 3.2). A proof can be seen in Appendix B.4. The essential idea of this regime is firstly letting the error accumulate for a τ period. Then, we estimate it by a KF ($\widehat{x}_{t,\tau}$), so, for each small step forecasting within τ horizon, we adjust the forecast by $\widehat{x}_{t,\tau} \cdot f(\tau)$, and transition coefficients in the Kalman filter F and Q:

$$\widehat{V}_{t+s}^{(n_i)} := \widehat{V}_{t+s}^{(n_i)} + F \cdot Q \cdot \widehat{x}_{t,\tau} \cdot f(\tau)$$

which becomes Eq. 3.12. Again, we remark the reason for translating $\widehat{x}_{t,\tau}$ to $\widehat{x}_{t+s,1}$ for $s = 0, 2, .., \tau - 1$ is better captures jump clustering features in a higher frequency. We will show in our empirical works that jump clustering features can diminish with the decrease in data frequency.

3.4.3 Simulation Studies on a Simple ASV Model

In this section, we perform simulations on a simple ASV model (the ASV model in Table 3.3) to study the resources of forecast errors and the effectiveness of KF. For simplicity, we do not consider the intraday periodicity in this section. The model reads:

$$Y_t = \mu + \sqrt{V_{t-1}}\epsilon_t^Y + \eta_t, \tag{3.26}$$

$$V_t = V_{t-1} + \kappa(\theta - V_{t-1}) + \sigma_V \sqrt{V_{t-1}} \epsilon_t^V \tag{3.27}$$

Interpretations of parameters in this model are the same as those in Eq. 3.1 and 3.2. Similarly, we assume a correlated Brownian motion ($corr(\epsilon_t^Y, \epsilon_t^V) = \rho$) to add the leverage effect in this model. In addition, we consider the MN by adding an additional noise (η_t) in the return process. Inspired by Li et al. (2021), we consider the following 4 different types of MN:

$$\eta_t = \begin{cases} \frac{1}{5} \cdot \sqrt{V_{t-1}} \cdot \phi_t & \text{(Gaussian noise)} \\ \frac{1}{5} \cdot \sqrt{V_{t-1}} \cdot \omega_t^B \sqrt{\frac{\upsilon-2}{\upsilon}} & (t\text{-distributed noise}) \\ \frac{1}{5} \cdot \sqrt{V_{t-1}} \cdot \omega_t^A & \text{(correlated Gaussian noise)} \\ \frac{1}{5} \cdot \sqrt{V_{t-1}} \cdot \left(\frac{1}{2}\omega_t^A + \frac{1}{2}\omega_t^B \sqrt{\frac{\upsilon-2}{\upsilon}} \right) & \text{(Gaussian-}t\text{ mixture noise)} \end{cases} \tag{3.28}$$

where ϕ_t is a random variable from the standard normal distribution $\mathcal{N}(0,1)$. ω_t^B is a random variable from a t-distribution with a degree of freedom υ. ω_t^A is a correlated Gaussian random variable defined as follow:

$$\omega_t^A = \phi_t + \sum_{j=1}^{\Lambda} \beta_j \phi_{t-j}, \quad \beta_j = \frac{d(1+d)\cdots(j-1+d)}{j!} \tag{3.29}$$

where Λ is a cut-off value, and $d \in (-0.5, 0.5)$. The correlated Gaussian-t mixture noise setting can capture some characteristics of MN in the stock market: 1) slowly-decaying auto-correlations; 2) volatility-dependence; 3) large bounce-backs in the transaction data (see Clinet & Potiron (2021) and Jacod et al. (2017)). Jacod et al. (2017) also show the ratio of MN's size to volatility is time-varying ranging from 0.1 to 0.7. We take a conservative estimate as $\frac{1}{5}$, but we remark that this is a trivial issue. We also tried MN with bigger sizes, and the effect of MN on the model can only become more evident.

We use this model as the data generating process (DGP), and set the parameters

in the DGP as follow[5]:

$$\begin{aligned} \mu = 0.3\Delta t \quad \kappa = 0.05\Delta t \quad \theta = 0.3\Delta t \quad \sigma_V = 0.2\Delta t \quad \rho = -0.3 \\ d = 0.3 \quad \Lambda = 100 \quad \upsilon = 2.5 \end{aligned} \tag{3.30}$$

where we take $\Delta t = \frac{1}{78}$ standing for 5-minute frequency.

We firstly conduct five simulations with different MN specifications, and simulate 10,000 days' 5-minute data for each simulation (780,000 data points). Then we estimate the model by Bayesian MCMC (with 100,000 iterations and a 50,000 burn-in period). Table 3.1 reports the posterior mean and standard deviation of parameters. As is shown in the table, some estimated parameters deviate from their actual values with the presence of MN. For example, κ inflates by nearly three times. Also, ρ becomes insignificant with MN. This step aims to show that estimated parameters can become inconsistent with the presence of MN.

[INSERT TABLE 3.1 ABOUT HERE]

Then, we show that forecasting with these inconsistent parameters can lead to the accumulation of forecast errors. Using the simulated data with correlated Gaussian-t mixture MN, we set the last 1000 days' data as the out-of-sample data. Then we estimate the model, and forecast variance. We firstly use estimated parameters under DGP with the Gaussian-t mixture noise (inconsistent parameters) to forecast variance. Then we use estimated parameters under DGP without noise (consistent parameters) to do the forecast as a comparison. Figure 3.2 plots the implied variance and actual variance. The figure shows that implied variance tracks the actual variance.

[INSERT FIGURE 3.2 ABOUT HERE]

Then, we correct errors in variance forecasts using the KF. Forecast results are plotted in Figure 3.3, 3.4 and 3.5. The top three graphs plot variance forecasts ($\widehat{V}_t$)

[5]We estimate the model using S&P 500 log returns (times 100 as a multiple) from 2012 to 2019, and retrieve these parameters.

against actual in each figure. The middle three graphs plot forecast errors of forecasts (red line, $u_{t,\tau}$ in Eq. 3.18), error forecasts by KF (green line, $\widehat{x}_{t,\tau}$ in Eq. 3.23) and forecast errors after the correction using KF (blue line). The bottom three graphs plot error-corrected forecasts. Figure 3.3 plots forecasts using consistent parameters, and Figure 3.4 and 3.5 plots forecasts using estimated consistent parameters. We consider multiple forecast horizons τ as we do in the empirical studies.

[INSERT FIGURE 3.3 ABOUT HERE]

[INSERT FIGURE 3.4 ABOUT HERE]

[INSERT FIGURE 3.5 ABOUT HERE]

As shown in Figure 3.3, when using consistent parameters in the forecasting, KF does not capture anything in forecast errors, and error forecasts (green line, $\widehat{x}_{t,\tau}$ in Eq. 3.23) are almost horizontal. Also, there are nearly no differences between forecasts with and without error correction. However, when using inconsistent parameters, forecast errors can be captured by KF, and the green line tracks forecast errors. In addition, we also show how forecast errors can accumulate with the increase in forecast horizons. Variance forecasts deviate further from actual variance when forecast horizons are greater than 24 (see Figure 3.4 and 3.5). Therefore, we show that 1) microstructure noise in the high-frequency data can lead to parameters being estimated inconsistently; 2) forecasting using these inconsistent parameters is subject to forecast errors which are auto-correlated and can accumulate over forecast horizons; 3) KF is useful in correcting these errors.

3.5 Empirical Application

We apply the model to 5-minute high-frequency data of an individual stock and a broad index. In this section, we introduce our dataset and in-sample estimations, and we will then report our out-of-sample forecasting performance along with evaluation methods.

3.5.1 Data

We retrieve 5-minute high-frequency data from Bloomberg for Apple Inc. (AAPL) and S&P 500 (SPX) from 3/1/2012 to 31/12/2019. We separate these eight years of data into 4,1,3 years for in-sample estimation, Kalman filter parameter estimation, and out-of-sample forecasting period. We remark that the one-year window is a pseudo out-of-sample period when we conduct the forecast, but only for the purpose of estimating parameters in the Kalman filter. We also examine 10-minute and 30-minute data to investigate the dynamics of jump clustering. The number of jumps detected in both data are reported in Table 3.2, clearly AAPL reports more return jumps than SPX does. For comparisons, we use four other stochastic volatility models as benchmarks described in Table 3.3.

[INSERT TABLE 3.2 ABOUT HERE]
[INSERT TABLE 3.3 ABOUT HERE]

3.5.2 In-Sample Estimation and model Fitness

We conduct non-parametric estimations on the periodic component $s_t = I_{tk}\hat{f}_k^{WSD}$, $k = 1, ..., M$ and jumps component $\{J_t^{Y+}, J_t^{Y-}, J_t^V\}$. Figure 3.6 plots f_k^{WSD} of AAPL and SPX with and without overnight returns, two graphs on the right-hand side are simply two graphs on the left-hand side without the first points. This figure shows that intraday periodicity appears an 'U-shape'. Additionally, Figure 3.7 and Figure 3.8 plots some variables in the return and variance processes. These figures show that there are much more return jumps than variance jumps. Also, AAPL has more return jumps than SPX does.

[INSERT FIGURE 3.6 ABOUT HERE]
[INSERT FIGURE 3.7 ABOUT HERE]
[INSERT FIGURE 3.8 ABOUT HERE]

Parameters estimated in the model are reported in Table 3.4, MHP parameters of SVIJ-MHP model are reported separately in Table 3.5. In assessing model fitness, we calculated the deviance information criterion (DIC) at the bottom of Table 3.4. The DIC is calculated by model log-likelihood penalised by model complexity and can be calculated simultaneously with MCMC estimation. A smaller value of DIC indicates a better fitness of the model. In addition, we also calculate the log-likelihood of the model and Bayes factor (likelihood of two models). We plot the cumulative log Bayes factor of 4 other models against SV model in Figure 3.9. As a result, SVIJ-MHP model reports the best fitness of models in terms of DIC and Bayes factor. Figure 3.9 also indicates the importance of including jumps in a model. From the figure, we can see log-likelihood of SVJ, SVIJ and SVIJ-MHP model are higher than that of ASV and SV model.

[INSERT TABLE 3.4 ABOUT HERE]

[INSERT TABLE 3.5 ABOUT HERE]

[INSERT FIGURE 3.9 ABOUT HERE]

3.5.3 High-Frequency Variance Forecasting

To show the usefulness of the Kalman filter in error corrections, we plot variance forecasts of SPX using 5-minute data before and after the correction by Kalman filter in Figure 3.10. We compare variance forecasts by using mean square error (MSE), forecast bias ($\frac{\widehat{V_{t,\tau}}-V_{t,\tau}}{\widehat{V_{t,\tau}}+V_{t,\tau}}$), Diebold Mariano (DM) test[6] (Diebold & Mariano 1995), and Mincer-Zarnowitz regressions ($V_{t,\tau} = \beta_0 + \beta_{mz}\widehat{V_{t,\tau}} + v_{t,\tau}$, (Mincer & Zarnowitz 1969)).

Remark 6 (Proxy for Actual Values of Variance) *Our proxy of true values of*

[6] For example, we have two time-series of forecast errors ($e_{t,1}$ and $e_{t,1}$) from two different models. We calculate Diebold-Mariano statistic as $\mathrm{DM} = \bar{s}/\left(\sqrt{(\gamma_0 + 2\sum_{k=1}^{h-1}\gamma_k)/n}\right)$, where $\bar{s} = \mathbb{E}(e_{t,1}^2 - e_{t,1}^2)$ and γ_k denotes the autocovariance of $e_{t,1}^2 - e_{t,1}^2$ at lag k. n denotes number of forecasts and we take $h = n^{\frac{1}{3}} + 1$.

variance is estimated using the full sample in the SVIJ-MHP model (model-implied variance). Note that we are aware of the suspicion of using model-implied variance as true values. We find difficulties in choosing a relatively 'perfect' proxy for high-frequency variance, and there is limited literature studying high-frequency variance forecasting. To robustify our results, we also use several different proxies. We follow Stroud & Johannes (2014), who use hourly RV as actual values in hourly variance forecasting. We also use implied variance by SV model as actual values. We additionally conduct simulation studies, and use simulated variance as actual values. We find results are similar, and all these additional results are confined to the appendix B.5. We also remark that this proxy problem only arises in high-frequency variance forecasting. Daily RV, VaR and ES forecasting in later sections do not have this issue. Their results also indicate the superiority of our model and the importance of considering jump clustering.

Table 3.6 reports the forecast performances of AAPL and SPX across different forecast horizon τ, forecasts by the SVIJ-MHP model provide the lowest MSE, smallest BIAS and highest R^2 in Mincer-Zaenowitz regressions, p-value of DM test also indicates that forecasts by SVIJ-MHP model are significantly different from those by other models. We also assess 10-minutes and 30-minutes data, and full results are provided in appendix B.5.

[INSERT FIGURE 3.10 ABOUT HERE]

[INSERT TABLE 3.6 ABOUT HERE]

In addition, to evaluate the effect of including jump clustering features, we also consider conditional performance by extracting h forecasts after different types of jumps. The purpose is to evaluate the forecasting contribution of jump clustering effect produced by different types of jumps. For example, Table 3.7 report conditional MSE of variance forecasts across different τ and h using AAPL 5-minute data, and Table 3.8 reports p-values of corresponding DM tests. Further, conditional MSE of using SPX data and associated DM test are provided in Table 3.9 and Table 3.10.

In general, the SVIJ-MHP model provides the best performance, especially using AAPL 5-minute data. For example, in 0.5-2 hours variance forecasts, MSE of the SVIJ-MHP model after negative return jumps is reduced by over 10% in AAPL data, while this figure is less than 3% in SPX data. However, results of DM tests suggest forecasts by SVIJ-MHP model are mostly different from those by other models except 2-3 hours forecasts after variance jumps using SPX 5-minute data.

[INSERT TABLE 3.7 ABOUT HERE]
[INSERT TABLE 3.8 ABOUT HERE]
[INSERT TABLE 3.9 ABOUT HERE]
[INSERT TABLE 3.10 ABOUT HERE]

In terms of 10,30-minutes data, we compare forecasts performance after return jumps in Table 3.11 and 3.12, and that after variance jumps in Table 3.13. We only provide DM test results here, and full results can be found in appendix B.5. We find that in SPX data, forecasts by SVIJ-MHP model become indifferent to those by other models with 10-minutes data. It still performs well in forecast half an hour variance after negative return jumps but generally becomes indifferent with other models in 30-minutes data after all types of jumps, while forecasts using AAPL data provide better performance in general. We also find that forecast performance after negative return jumps tends to be better than that after positive return jumps and variance jumps. Furthermore, we find advantages of using higher frequency data. For example, in forecasting 2-hour variance after negative return jumps, forecasts using 5-minute data outperform those using 10 and 30-minutes data. This suggests that advantages of jump clustering diminishes with the increase in data frequency.

[INSERT TABLE 3.11 ABOUT HERE]
[INSERT TABLE 3.12 ABOUT HERE]
[INSERT TABLE 3.13 ABOUT HERE]

These results suggest 1) Using AAPL data, which has more jumps in data, provides generally better variance forecasts performance than SPX when using SVIJ-MHP model; 2) negative return jump clustering helps more in forecasting variance; 3) higher-frequency data provides better forecasting performance; 4) forecast performance is comparatively better with forecasting horizons of less than 2 hours.

Our results are also in line with those in Chen et al. (2021), which show that the cluster lengths of high-frequency jumps are around 2-3 hours, negative jump clusters last longer and variance jumps are comparatively short-lived. Our results also indicate that contributions of including jump clustering to variance forecasting are greater with forecasting horizons less than 2 hours. This provides implications to market practitioners, especially high-frequency traders, that market risks can be raised up by jumps, especially 2 hours after spotting a jump.

3.5.4 Daily Realised Volatility Forecasting

Realised volatility (RV) has become a very popular volatility estimator since proposed by Andersen & Bollerslev (1997). Forecasting on RV has also attracted many studies. In our RV forecasting, we benchmark a heterogeneous autoregressive (HAR-RV) model, and its extension with continuous jump (HAR-RV-CJ) and signed jump (HAR-RV-SJ) (see Corsi (2009);Andersen et al. (2007);Patton & Sheppard (2015)). We also consider realised GARCH family models, including realised-GARCH model, its log-linear form and realised-EGARCH model (see Hansen et al. (2012);Hansen & Huang (2016)). In addition, we compare our model with models without error corrections, and 'no-change' forecasts as a baseline benchmark, which simply takes the previous day's RV as forecasts.

We obtain point estimates of daily RV by calculating the posterior mean of the simulated posterior distribution. Figure 3.11 plots the upper and lower 5% quantile of the RV distribution and daily RV of SPX in the out-of-sample period. Table 3.14 reports MSE, forecast bias and p-values of DM test on SVIJ-MHP forecasts against others. The R^2 in the table denotes goodness-of-fit in the Mincer-Zarnowitz regres-

sion (regressing true value with forecasts). In general SVIJ-MHP model outperform all other models with the lowest MSE and highest R^2. Results of the DM test also suggest SVIJ-MHP forecasts are significantly different from other forecasts. We also find that in SPX data, SVIJ-MHP model does not outperform other stochastic volatility models as much as it does in AAPL data. However, the DM test suggests their forecasts are indifferent. We also highlight that RV forecast results using a model without error correction by KF also report a reasonable R^2, which shows good fitness. However, they also report very high MSE and BIAS.

[INSERT FIGURE 3.11 ABOUT HERE]
[INSERT TABLE 3.14 ABOUT HERE]

We further run a regression to examine the significance of SVIJ-MHP model forecasts. We further regress forecast errors of SVIJ-MHP model with forecasts by other models:

$$RV_t - \widehat{RV}_{t,SVIJ-MHP} = b_0 + b_1 \widehat{RV}_{t,other}, \tag{3.31}$$

and Table 3.15 reports the results, all coefficients are tested to be insignificant suggesting other models' forecasts do not help explain forecast errors of SVIJ-MHP model.

[INSERT TABLE 3.15 ABOUT HERE]

Overall, in RV forecasting, SVIJ-MHP outperforms HAR-class models and realised-GARCH class models. It also outperforms other benchmark stochastic volatility models, which shows the predictability of our model. Moreover, comparing forecast results from SVIJ-MHP and SVIJ model, the improvement in AAPL data can indicate the advantages brought by the inclusion of the MHP processes, hence, jump clustering features.

3.5.5 VaR and ES Forecasting

As two popular risk measures, VaR and ES can provide estimates of potential losses for banks and financial institutions. Our estimates of VaR and ES are based on posterior distributions of simulated return. To evaluate out-of-sample forecasting performance, we adopt a loss function proposed by Fissler & Ziegel (2016) to calculate the loss of VaR and ES jointly:

$$\mathcal{L}_{FZ}(Y_t, VaR, ES, \alpha_{ve}) = -\frac{1}{\alpha_{ve} ES}\mathbb{1}\{Y_t < VaR\}(VaR - Y_t) + \frac{VaR}{ES} + log(-ES) - 1 \tag{3.32}$$

where α_{ve} denotes significant levels of VaR and ES. Table 3.16 and 3.18 report expected losses of VaR and ES forecasts using AAPL and SPX 5-minute data, and the expected loss of forecasts by the SV model is scaled to 1. We also conduct a DM test on SVIJ-MHP forecasts against other forecasts. Results suggest that SVIJ-MHP model reports the lowest expected loss. Additionally, the model performs better under higher significant levels than those under lower significant levels. The inclusion of the MHP process can reduce the expected loss by up to 15% compared to the expected loss of SVIJ model forecasts. Further, according to DM test results, SVIJ-MHP forecasts are significantly different from forecasts by other models.

[INSERT TABLE 3.16 ABOUT HERE]

[INSERT TABLE 3.18 ABOUT HERE]

In backtesting VaR and ES, we firstly adopt a Dynamic Quantile (DQ) test proposed by Engle & Manganelli (2004). This test focus on whether violations of VaR ($Y_t < VaR_t$) are serially independent distributed or correlated with VaR estimator. Defining a 'Hit' variable as $Hit_t = \mathbb{1}\{Y_t < VaR_t\} - \alpha_{ve}$, consider the following regression:

$$Hit_t = c_0 + c_1 Hit_{t-1} + c_2 VaR_{t-1} + u_{v,t} \tag{3.33}$$

We are interested in the coefficient estimates $\{c_0, c_1, c_2\}$ are jointly equal to 0, the DQ test gives a test statistic which is asymptotically chi-squared distribution ($\mathcal{X}^2(3)$).

We also backtest ES forecasts by a Dynamic Expected Shortfall (DES) regression proposed by Patton et al. (2019):

$$\lambda^s_{e,t} = d_0 + d_1\lambda^s_{e,t-1} + d_2 ES_{t-1} + u_{e,t}, \tag{3.34}$$

where $\lambda^s_{e,t} = \frac{1}{\alpha_{ve}}\mathbb{1}\{Y_t < VaR_t\}\frac{Y_t}{ES_t} - 1$, similarly, we looks at whether $\{d_0, d_1, d_2\}$ are jointly equal to 0. Table 3.17 and 3.19 reports the coverage ratio of VaR forecasts, p-value of DQ and DES tests. We can see from the results that SVIJ-MHP model reports the lowest coverage ratio in all cases, in the sense that SVIJ-MHP provides lower forecast values of VaR when including jump clustering features. Regarding DQ and DES tests, SVIJ-MHP provides the least number of tests being rejected across all forecast horizons and significance levels, meaning that SVIJ-MHP forecasts of VaR and ES are least likely serially correlated and least likely correlated with VaR and ES estimates. However, we also notice that DQ and DES tests of SPX 5-minute data perform differently compared to those of AAPL 5-minute data. DQ and DES tests in SPX data report a much lower number of tests being rejected.

[INSERT TABLE 3.17 ABOUT HERE]

[INSERT TABLE 3.19 ABOUT HERE]

Similar to variance forecasts, we also consider 1) conditional performance after different types of jumps; 2) performance with different data frequencies (10,30-minutes). We found SVIJ-MHP outperforms other models in forecasting VaR and ES with forecasting horizons of less than 1 hour in terms of DM tests. Also, performance after negative return jumps is better than that after positive return jumps and variance jumps. Further, performance using AAPL data is better than that using SPX data. Details of evaluation on condition performance and performance using data with different frequencies are confined to the appendix B.5.

3.6 Conclusion

In this chapter, we propose a high-frequency forecasting framework based on a high-frequency stochastic volatility model featuring jump clustering. We apply a Bayesian MCMC algorithm in the in-sample estimation of static parameters. In out-of-sample forecasting, we use a particle filter to estimate latent states at each time point and forward simulate distributions of variance, returns and squared returns. We obtain point forecasts by taking the posterior mean of distributions. We find MN in the high-frequency data can affect models' estimated parameters, and we further apply a Kalman filter to correct the forecast errors due to inconsistent parameters. We conduct a simulation study to show the accumulation of forecast errors due to the inconsistent parameters due to MN and the effectiveness of error correction by KF. These are our novel contributions and to our knowledge, we are the first to do so.

In our empirical research, we apply 5-minute data of an individual index (AAPL) and a broad Index (SPX) and consider forecasting on variance, realised volatility and two other risk measures, VaR and ES. Our results suggest that including jump clustering features in forecasting models, especially forecasting on high-frequency variance and risk measures. We also find strong evidence that negative return jumps contribute more to forecasting performance than other types of jumps do. In addition, we show the importance of correcting errors when integrating forward simulated forecasts under high-frequency data.

We also examine conditional performance by considering forecast performance h period after occurrences of different types of jumps across different forecasting horizons. We find forecasts by our model outperform and are significantly different from forecasts by other models, especially when the forecasting horizon is less than 2 hours. Additionally, this advantage of including jump clustering tends to become less significant with decreasing data frequency. Further, we find an individual stock data containing more jumps provides better performance than a broad index data.

We also highlight that although jump clustering mostly happens intradaily, it

is also necessary to consider it in forecasting on daily measures. For example, our results suggest including jump clustering help improve the performance of realised volatility forecasting in terms of lower MSE and higher R^2 in Mincer-Zarnowitz regressions against two popular classes of realised volatility forecasting models (HAR and realised GARCH).

Overall, in forecasting VaR and ES, we find that expected losses of VaR and ES can be reduced up to 15% by including jump clustering. Furthermore, we find a similar pattern in variance forecasting that conditional forecast performance is better after negative return jumps with forecast horizons of less than 2 hours. We further backtest the dependence of VaR and ES violations by adopting DQ and DES tests, and we find our model significantly reduce the dependency of violations.

Volatility and VaR/ES forecasting are important for financial applications. With the increasing popularity of high-frequency trading and emphasis on financial market jumps, our studies are particularly useful for high-frequency traders and practitioners who are interested in intraday dynamics and high-frequency forecasting.

B Appendices for Chapter 3

B.1 Bayesian MCMC Algorithm and Specification of Priors

To obtain the joint posterior distribution in Equation 3.10, we adopt a Bayesian MCMC algorithm that randomly samples from a set of conditional posterior distributions. For $i = 1...300,000$, our sampling plan is as follow:

Algorithm 1 Bayesian MCMC Algorithm

For $i = 1 : 300,000$:

1: Sample k static parameters:

Draw $\Theta_1^{(i)}$ from $p\left(\Theta_1^{(i)}|Y_t, \Theta_2^{(i-1)}, \Theta_3^{(i-1)}, ..., \Theta_k^{(i-1)}, \xi_t^{Y+(i-1)}, \xi_t^{Y-(i-1)}, \xi_t^{V(i-1)}, \boldsymbol{\vartheta}^{(i-1)}\right)$,

$\vdots$

Draw $\Theta_k^{(i)}$ from $p\left(\Theta_k^{(i)}|Y_t, \Theta_1^{(i-1)}, \Theta_2^{(i-1)}, ..., \Theta_{k-1}^{(i-1)}, \xi^{Y+(i-1)}, \xi^{Y-(i-1)}, \xi^{V(i-1)}, \boldsymbol{\vartheta}^{(i-1)}\right)$

.

2: Sample return and variance jump sizes

for $t = 1, 2, ..., T$:

Draw $\xi_t^{Y+(i)}$ from $p\left(\xi_t^{Y+(i)}|Y_t, \Theta^{(i)}, \xi_t^{Y-(i-1)}, \xi_t^{V(i-1)}, \boldsymbol{\vartheta}^{(i-1)}\right)$,

Draw $\xi_t^{Y-(i)}$ from $p\left(\xi_t^{Y-(i)}|Y_t, \Theta^{(i)}, \xi_t^{Y+(i)}, \xi_t^{V(i-1)}, \boldsymbol{\vartheta}^{(i-1)}\right)$,

Draw $\xi_t^{V(i)}$ from $p\left(\xi_t^{V(i)}|Y_t, \Theta^{(i)}, \xi_t^{Y+(i)}, \xi_t^{Y-(i-1)}, \boldsymbol{\vartheta}^{(i-1)}\right)$.

3: Sample variance

for t = 1,2, ... , T:

Draw V_t from $p\left(V_t^{(i)}|Y_t, \Theta^{(i)}, \xi^{P+(i)}, \xi^{P-(i)}, \xi^{V(i)}, \boldsymbol{\vartheta}^{(i-1)}\right)$,

4: Sample the branching coefficient matrix

for $t = 1, 2, ..., T$:

Draw $\boldsymbol{\vartheta}^{(i)}$ from $p\left(\boldsymbol{\vartheta}^{(i)}|Y_t, \Theta^{(i)}, \xi_t^{Y+(i)}, \xi_t^{Y-(i)}, \xi_t^{V(i)}\right)$

We ran our MCMC for 300,000 iterations and took the first 50,000 as a burn-in period. For those conditional posterior distributions whose corresponding conjugate priors are available, we use Gibbs sampling methods, for those posteriors that are not available in closed form, we adopt the Metropolis-Hastings sampling method to approximate posteriors. Metropolis-Hastings method draws samples from a proposal distribution and a uniform-distributed random number. Then it decides whether the proposal draw should be accepted or rejected. Details of posterior distributions are provided by Stroud & Johannes (2014) and Rasmussen (2013). Our specification of priors are as follow:

[INSERT TABLE 3.20 ABOUT HERE]

B.2 Particle Filter

Our particle filter algorithm follows Pitt & Shephard (1999) and Stroud & Johannes (2014), Creal (2012) also gives a review on particle filter methods. Our ultimate goal of using a particle filter is to sample from $p\left(\mathcal{Z}_t|Y_t,\widehat{\Theta}\right)$ to obtain latent state at each time t where $\mathcal{Z}_t=\{V_t,\xi_t^{P+},\xi_t^{P-},\xi_t^V,\boldsymbol{\vartheta}\}$. We adopt a sequential importance sampling with re-sampling scheme:

Algorithm 2 Sequential Importance Sampling with Re-sampling

At $t=0$, for $i=1,...,N$

Draw $\mathcal{Z}_0^{(i)}\sim g_0(\mathcal{Z}_0)$, set $\omega_0^{(i)}=\frac{p(\mathcal{Z}_0^{(i)})}{g_0(\mathcal{Z}_0^{(i)})}$

For $t=1,...,T$:

1: For $i=1,...,N$, draw $\mathcal{Z}_t^{(i)}\sim g_t\left(\mathcal{Z}_t|\mathcal{Z}_{t-1},Y_t,\widehat{\Theta}\right)$.

2: Compute importance weights $\omega_t^{(i)}\propto\omega_{t-1}^{(i)}\frac{p\left(Y_t|\mathcal{Z}_t^{(i)},\widehat{\Theta}\right)p\left(\mathcal{Z}_t^{(i)}|\mathcal{Z}_{t-1}^{(i)},\widehat{\Theta}\right)}{g_t\left(\mathcal{Z}_t^{(i)}|\mathcal{Z}_{t-1}^{(i)},Y_t,\widehat{\Theta}\right)}$.

3: Normalise importance weights $\widehat{\omega}_t^{(i)}=\frac{\omega_t^{(i)}}{\sum_{j=1}^N\omega_t^{(j)}}$.

4: Re-sample N particles by $\{\widehat{\omega}_t^{(i)}\}_{i=1}^N$ and reset $\omega_t^{(i)}=\frac{1}{N}$.

More details and explanations about the algorithm can be found in Creal (2012).

B.3 Kalman Filter

Our goal of using a Kalman filter is to recursively update and predict the continuous component of forecast errors $x_{t,\tau},x_{t+\tau,\tau},x_{t+2\tau,\tau},...$ given forecast errors $u_{t,\tau},u_{t+\tau,\tau},u_{t+2\tau,\tau},....$ The algorithm is as follow:

Algorithm 3 Kalman Filter

At time t, when the forecast error $u_{t,\tau}$ comes in, we update $x_{t,\tau}$:

1: Compute Kalman gain: $K_{t,\tau} = \frac{F \cdot p_{t-\tau,\tau}}{F \cdot p_{t-\tau,\tau} + \sigma_\eta^2}$

2: Update $x_{t,\tau}$ estimates: $\widehat{x}_{t,\tau} = Q \cdot \widehat{x}_{t-\tau,\tau} + K_{t,\tau}(u_{t,\tau} - F \cdot \widehat{x}_{t-\tau,\tau})$

3: Update estimates uncertainty: $p_{t,\tau} = p_{t-\tau,\tau}(1 - K_{t,\tau})$

Then we predict $x_{t+\tau,\tau}$:

1: Predict $x_{t+\tau,\tau}$: $\widehat{x}_{t+\tau,\tau} = Q \cdot \widehat{x}_{t,\tau}$

2: Predict $x_{t+\tau,\tau}$ uncertainty: $p_{t+\tau,\tau} = Q \cdot p_{t,\tau} + \sigma_\delta^2$

B.4 Proof of Proposition 1

Given estimated forecast error at time t over forecast horizon τ as $\widehat{x}_{t,\tau}$, assume it spreads evenly across each step over τ as $\widehat{x}_{t,1} = \widehat{x}_{t+1,1} = \cdots = \widehat{x}_{t+\tau-1,1}$. Considering $V_{t,1}$ and $V_{t+1,1}$, using the variance process in Eq. 3.2, we obtain:

$$V_{t+1,1} = V_{t,1} + \kappa(\theta - V_{t,1}) + \sigma_V \sqrt{V_{t,1}} \epsilon_t^V + \xi_t^V J_t^V \tag{IB.4.35}$$

Here we neglect the impact of the jump component. We adjust $V_{t,1}$ by $\widehat{x}_{t,1}$, $V_{t,1} := V_{t,1} + \widehat{x}_{t,1}$. Therefore, the increment adding to $V_{t+1,1}$ by doing this step is equal to $\widehat{x}_{t,1} \cdot (1 - \kappa)$. Similarly, we can obtain the contribution of adding $\widehat{x}_{t+s,1}$, $s = 0, 1, ..., \tau - 1$ at each step to the total variance forecasts $V_{t,\tau}$ as $\sum_{i=0}^{\tau-s}(1-\kappa)^i \cdot \widehat{x}_{t,1}$. Therefore, we can obtain the following equation:

$$\widehat{x}_{t,\tau} = \sum_{s=0}^{\tau-1} \sum_{i=0}^{\tau-s} (1-\kappa)^i \cdot \widehat{x}_{t+s,1} \tag{IB.4.36}$$

Denoting $f(\tau) = \sum_{s=1}^{\tau-1} \sum_{i=0}^{\tau-s} (1-\kappa)^i$, we can calculate:

$$f(\tau)(1-\kappa) - f(\tau) = \sum_{i=1}^{\tau} (1-\kappa)^i - \tau \tag{IB.4.37}$$

which becomes a geometry sequence, and we can solve the equation as follow:

$$\widehat{x}_{t,1} = \widehat{x}_{t,\tau} \cdot f(\tau), \quad f(\tau) = \kappa \left[1 + \tau - \frac{1 - (1-\kappa)^{\tau+1}}{\kappa} \right]^{-1} \tag{IB.4.38}$$

B.5 Supplementary Results

In this section, we describe the results of variance forecasts using realised variance and SV-model-implied variance as actual values. We also conduct some simulation studies, see a brief introduction in Section 1. We also look at additional forecast evaluation results using different data frequencies.

In our high-frequency variance forecasting (Section 4.3), we used model-implied variance as actual variance values. We first estimated variance using full sample, and split it into in-sample and out-of-sample period to do further estimation and forecasting. Clearly, one may argue using model-implied variance as actual values is suspicious, therefore, in this appendix, we show results of high-frequency forecasting using realised variance (RV) as actual values and also results using simulated data to robustify our results.

We firstly follow Stroud & Johannes (2014) who use 5-minute data to calculate hourly RV as actual values of variance. So, for variance at time t with τ forecast horizon, we denote:

$$V_{t,\tau} = RV_{t,\tau} = \sum_{s=1}^{\tau} (Y_{t+s})^2$$

where Y_t denotes return at time t. We select forecast horizon from 30 minutes to 3 hours $\tau \in \{6, 12, 24, 36\}$ since daily RV ($\tau = 78$) is evaluated in Section 4.4. However, we are using 5-minutes data to calculate RV, it is common to using it to calculate daily RV, we have to note here that using it to calculate high-frequency RV may also be suspicious.

In addition, we also conduct simulation studies and use simulated variance as actual values to compare forecasts. Data generating process (DGP) is using the model specified in Section 2, and parameters in DGP is set at parameters estimated using in-sample SPX data (see Section 4.2). However, using simulated data is also subjected to limitations, which may highly depend on DGP.

Furthermore, we use the implied variance (estimate the variance using full sample) by SV model instead of that by SVIJ-MHP model as we did in the main chapter.

As a consequence, results using RV as actual values and results in simulation studies are similar to those we report in the main text (using model-implied variance as actual values).

3.7 Figures and Tables for Chapter 3

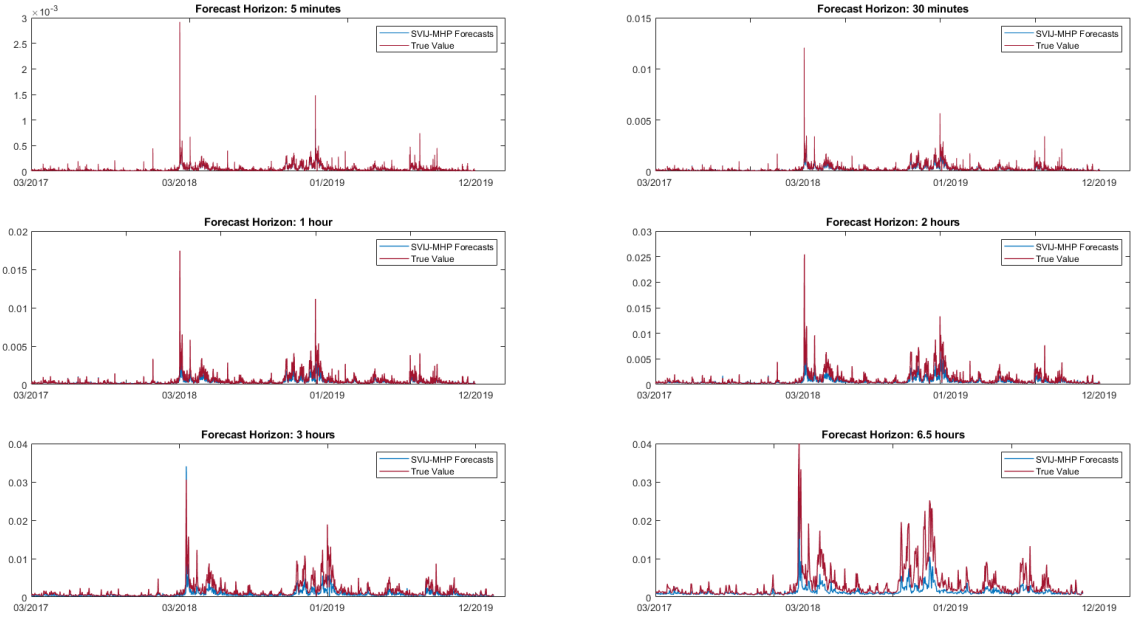


Figure 3.1: S&P 500 Index variance forecasts by SVIJ-MHP model and true values with different forecast horizons

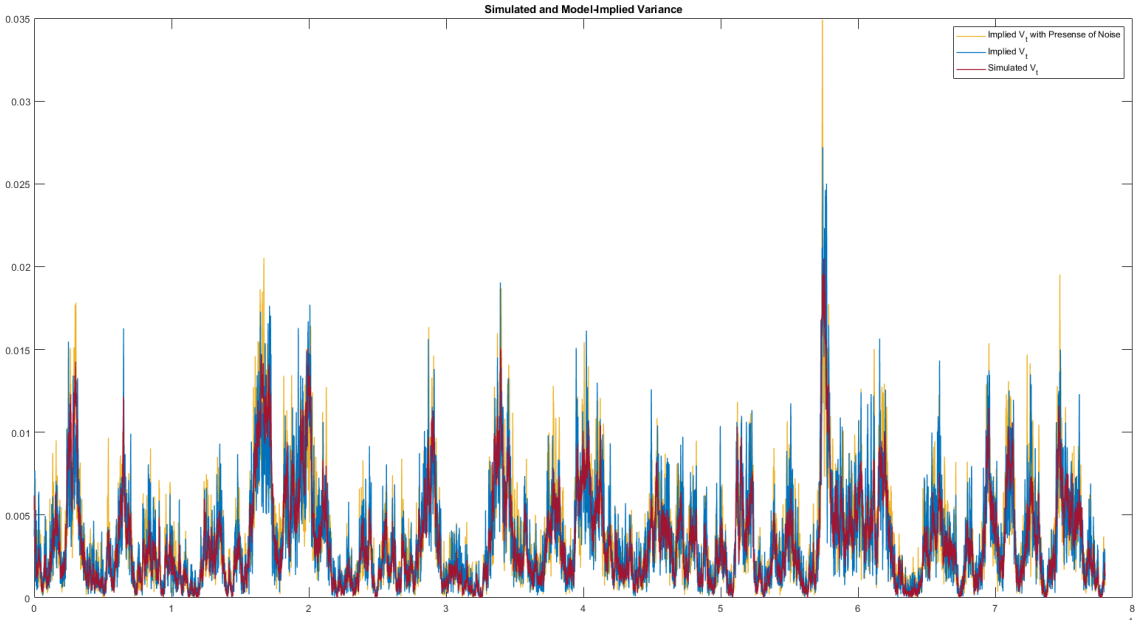


Figure 3.2: Implied Variance of SV model using Simulated Data with Different Frequencies

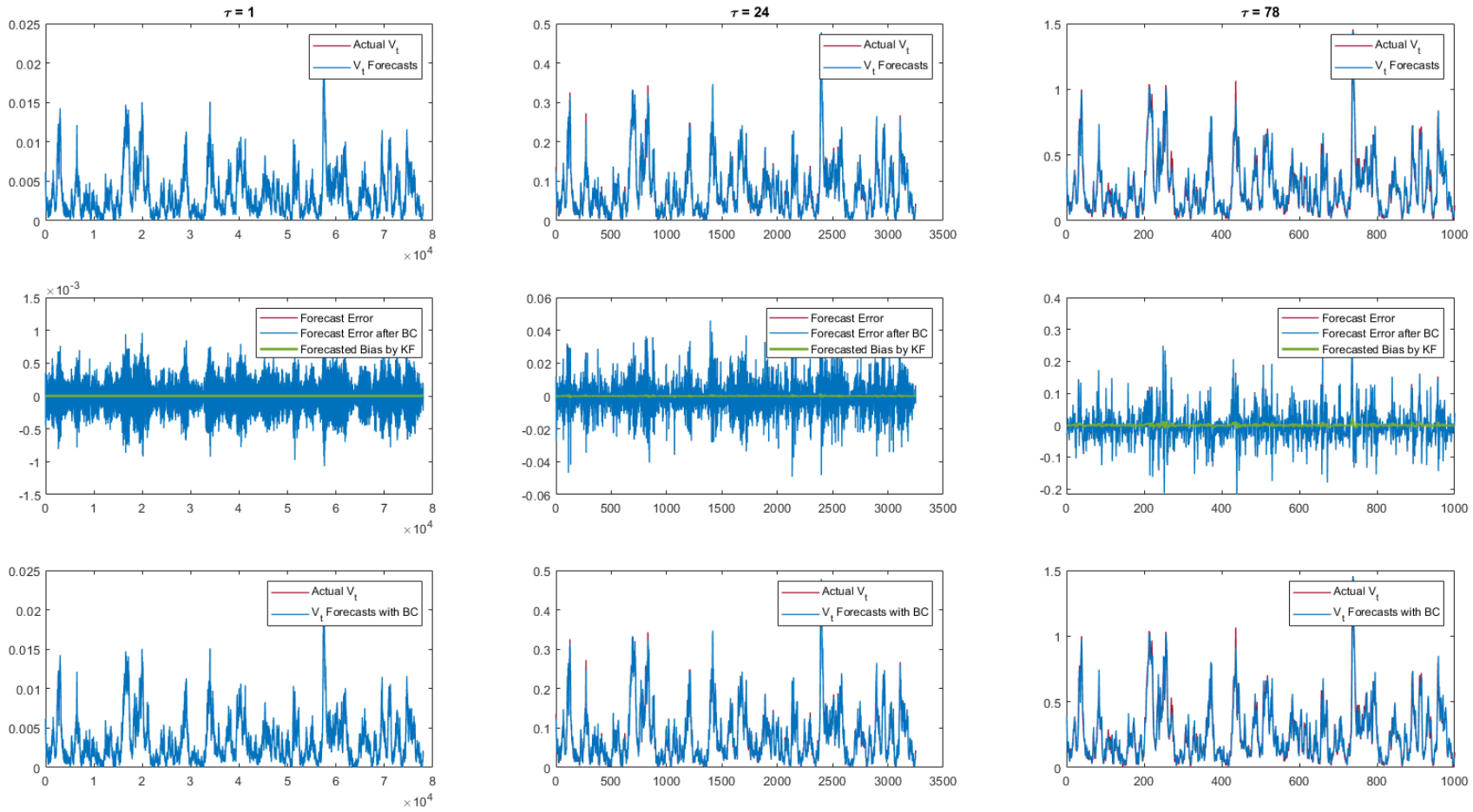


Figure 3.3: Simulated Volatility Forecasting and KF Correction with Consistent Parameters

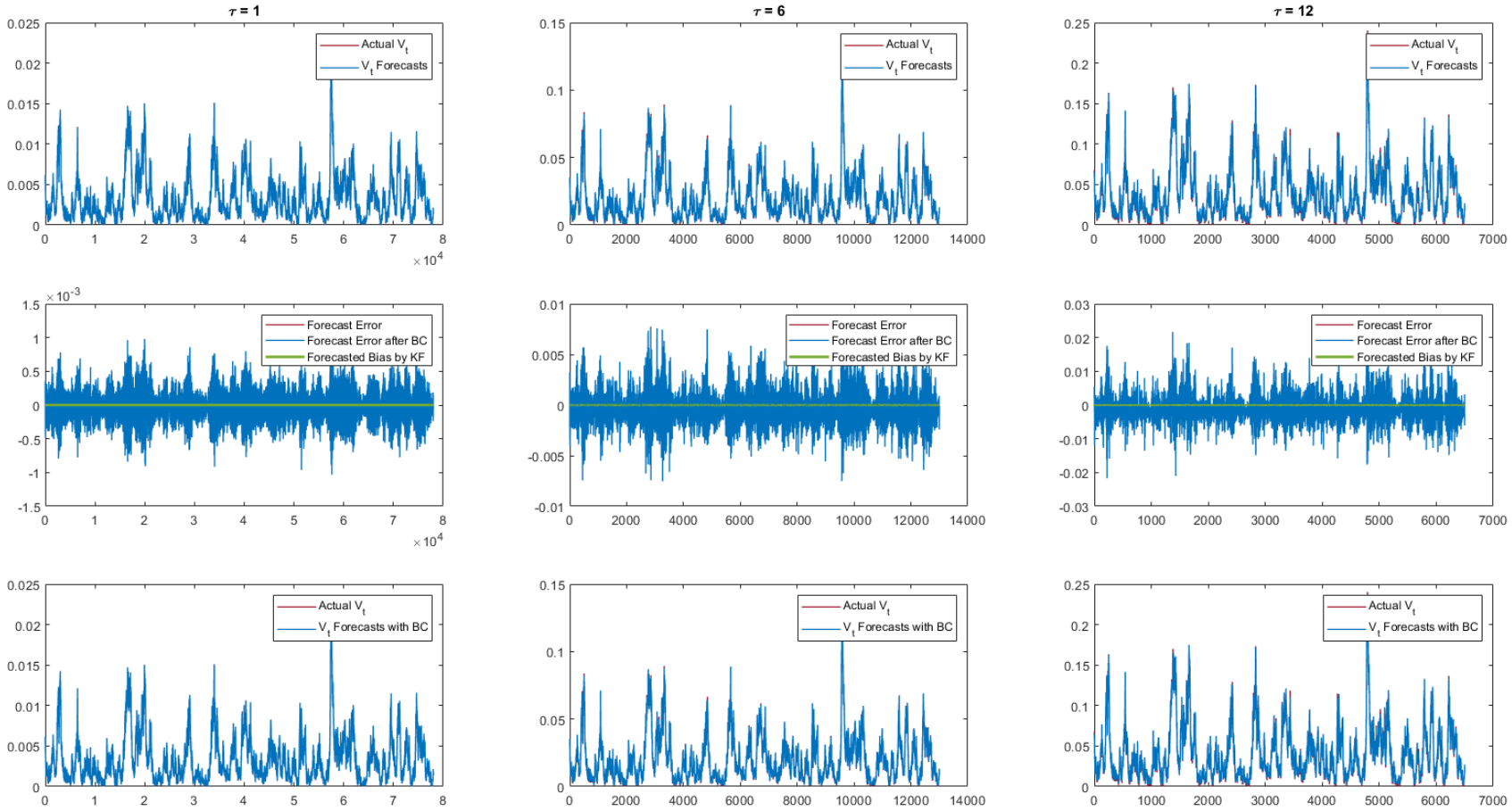


Figure 3.4: Simulated Volatility Forecasting and KF Corrections with Inconsistent Parameters

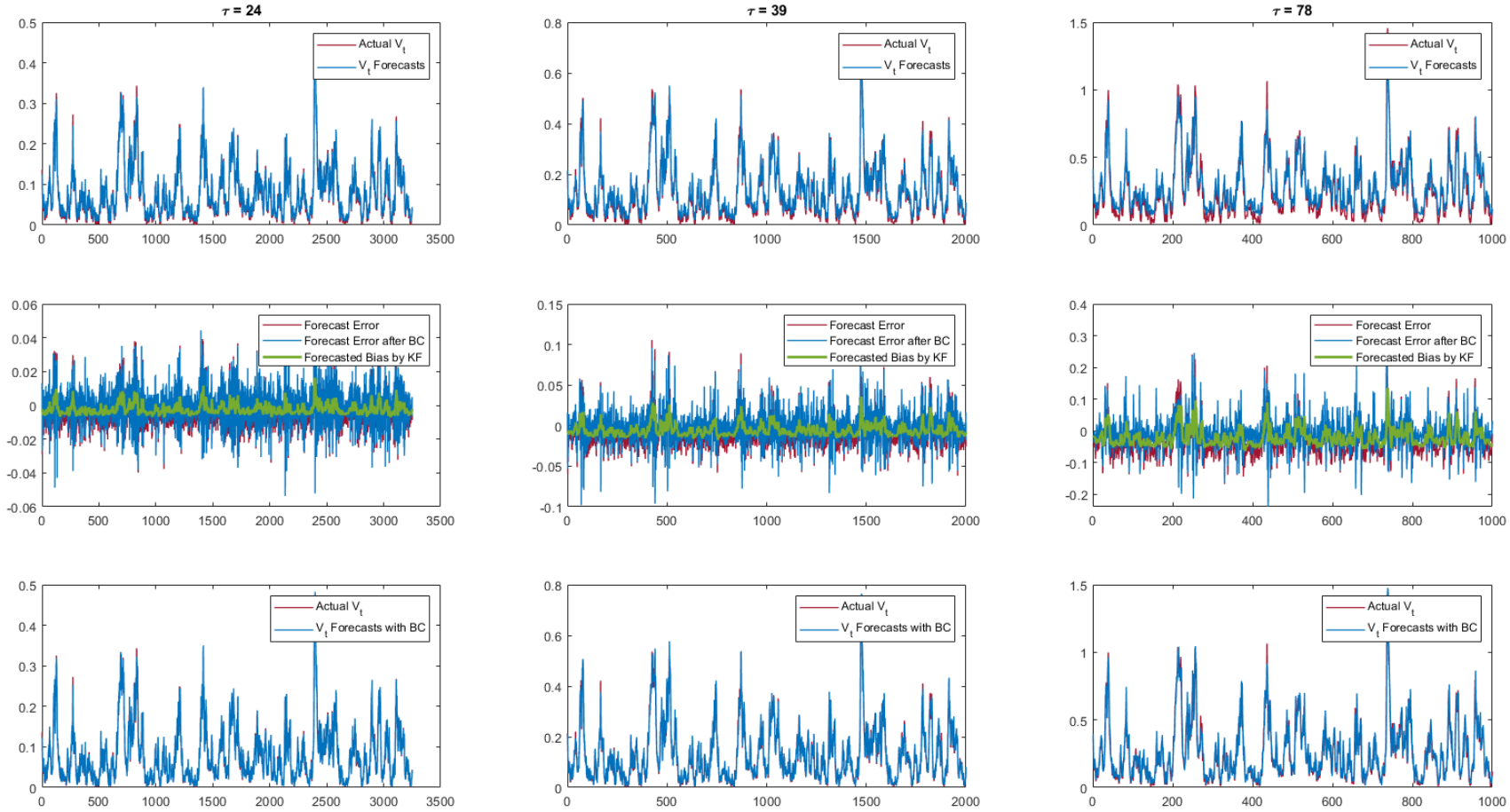


Figure 3.5: Simulated Volatility Forecasting and KF Corrections with Inconsistent Parameters (2)

Table 3.1: Estimated Parameters in Simulation Studies

	μ	κ	θ	σ_V	ρ	μ	κ	θ	σ_V	ρ
TRUE	0.3	0.1	0.3	0.2	-0.3	0.3	0.1	0.3	0.2	-0.3
without Noise						Presence of Correlated Gaussian Noise				
Posterior mean	0.29	0.094	0.32	0.23	-0.192	0.29	0.343	0.41	0.22	-0.006
Posterior std. dev.	0.01	0.016	0.01	0.02	0.033	0.01	0.007	0.01	0.02	0.003
Presence of Gaussian Noise						Correlated Gaussian-t Mixture Noise				
Posterior mean	0.31	0.239	0.36	0.18	-0.001	0.32	0.338	0.38	0.20	0.005
Posterior std. dev.	0.01	0.039	0.01	0.03	0.003	0.01	0.009	0.01	0.02	0.008
Presence of t-distribution noise										
Posterior mean	0.31	0.370	0.43	0.25	0.019					
Posterior std. dev.	0.01	0.008	0.02	0.01	0.007					

Notes: this table presents posterior mean and standard deviation of estimated parameters. We simulate 780,000 data points for each simulation (5 in total) with different specifications of microstructure noise.

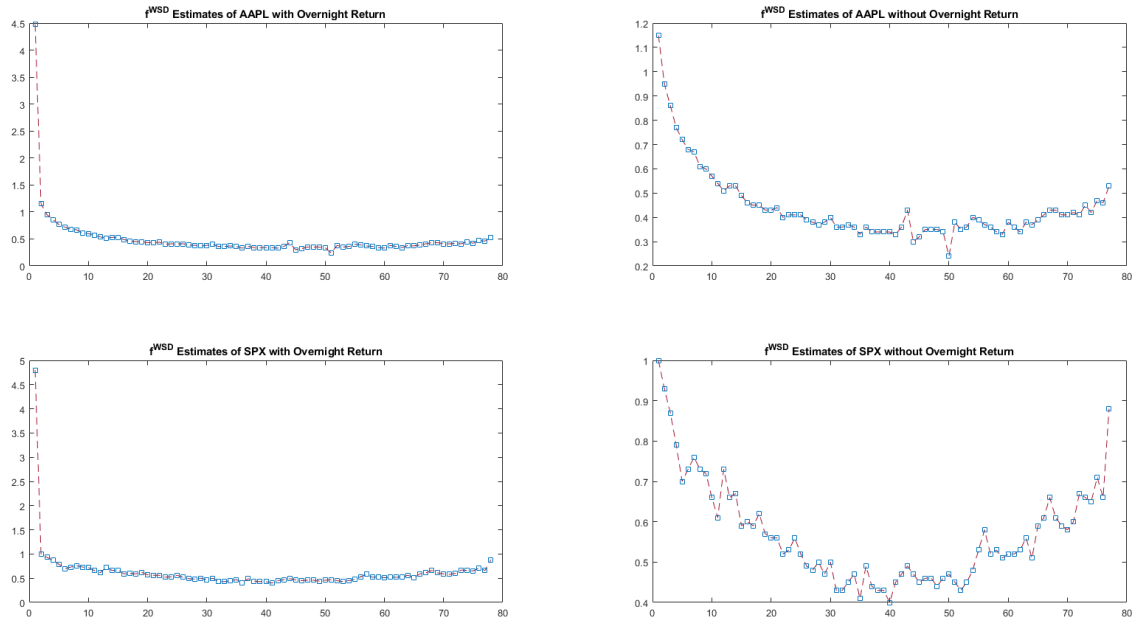


Figure 3.6: Periodic Components of AAPL and SPX with and without Overnight Return Estimated Using Data from 2012 to 2015 (in-sample)

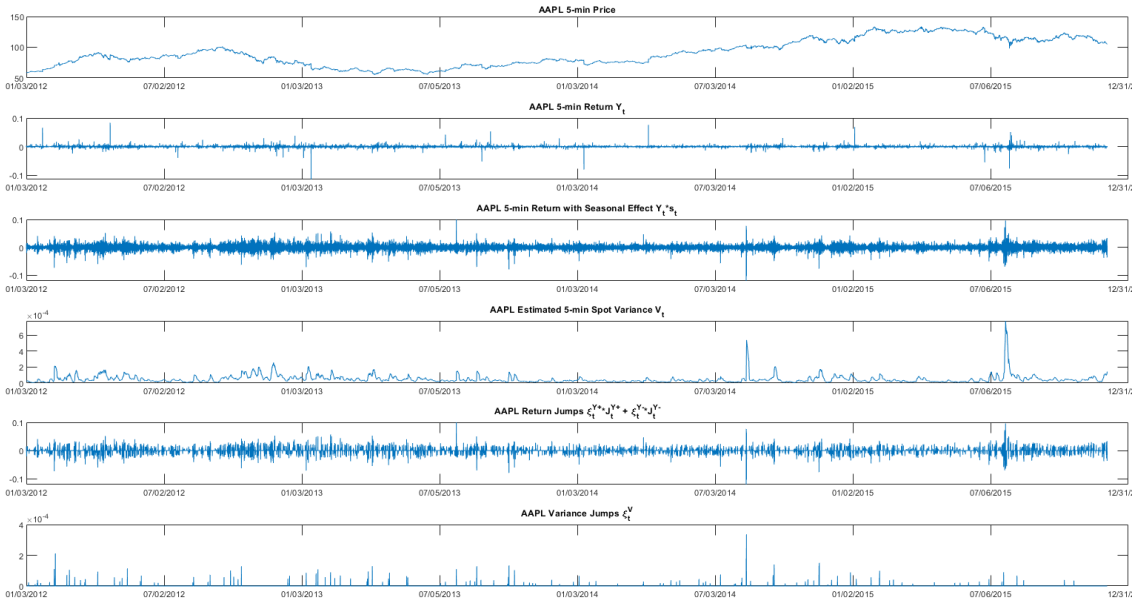


Figure 3.7: 5-min price, returns, estimated spot variance, return jumps, and variance jumps of AAPL

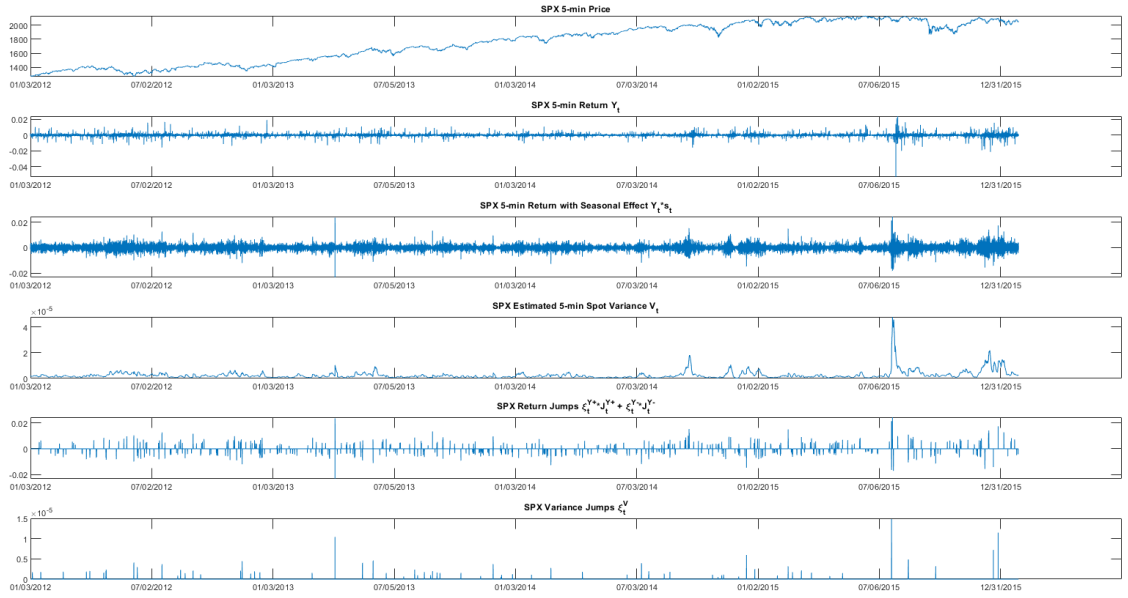


Figure 3.8: 5-min price, returns, estimated spot variance, return jumps, and variance jumps of SPX

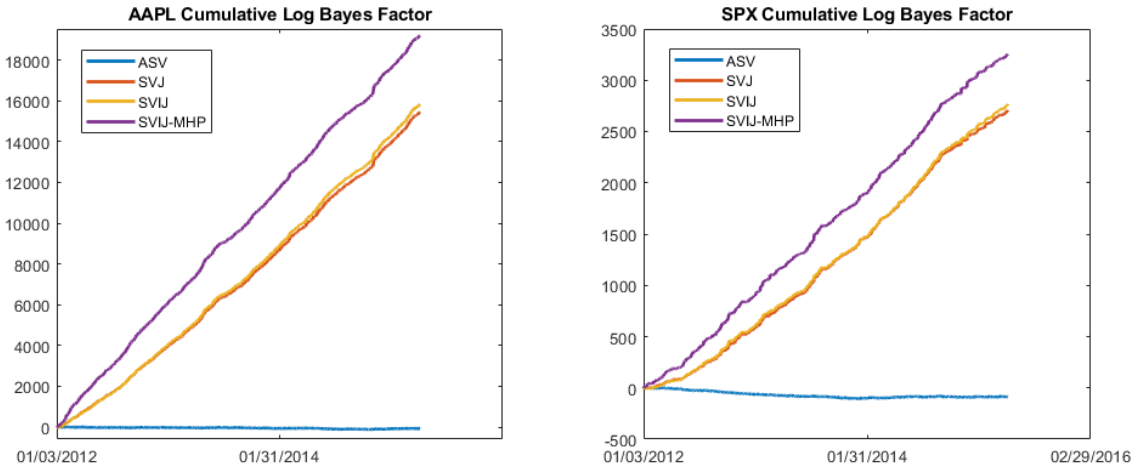
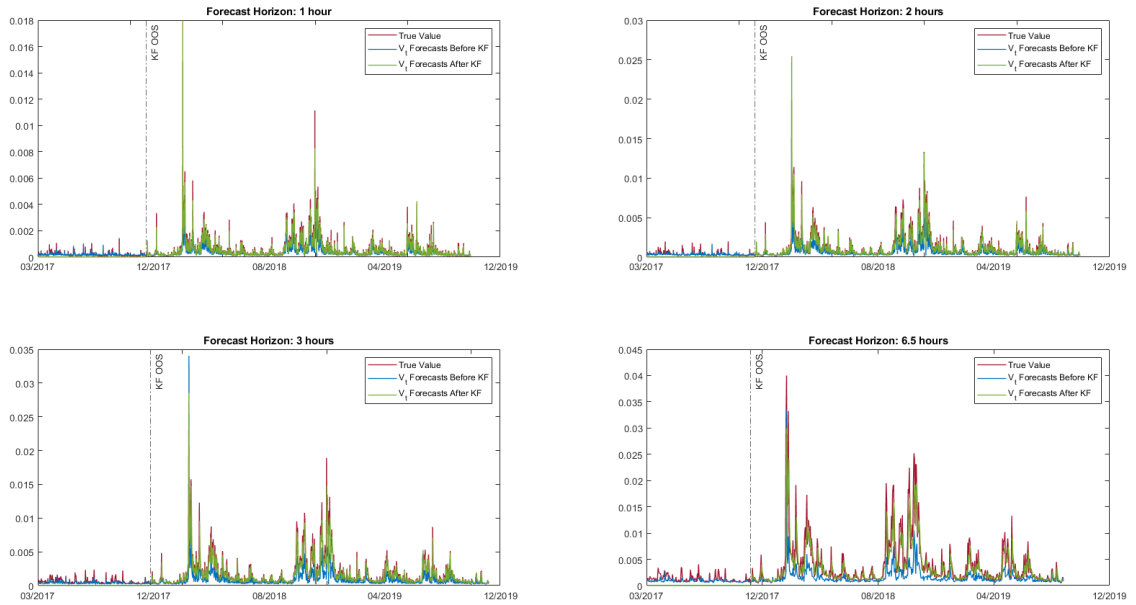


Figure 3.9: Cumulative Bayes Factor of AAPL and SPX



**The dash line 'KF OOS' marks the point before which parameters in Kalman Filter (KF) are estimated.*

Figure 3.10: SPX Variance Forecasts with KF across Different Forecasting Horizons

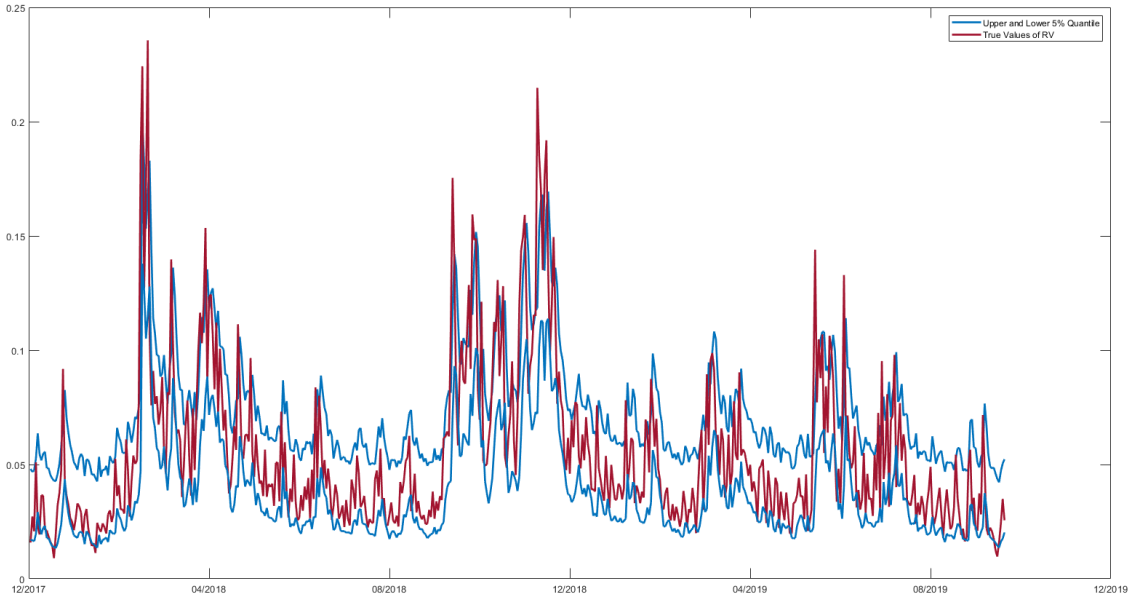


Figure 3.11: 5% and 95% Quantile of RV Posterior Distribution Using SPX Data

Table 3.2: Number of Jumps Detected in Samples

	AAPL		SPX	
	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample
Positive Return Jumps	909	756	337	165
Negative Return Jumps	953	904	473	298
Variance Jumps	177	153	162	112

Table 3.3: Stochastic Volatility Models Considered

	Leverage Effect	Return Jumps	Variance Jumps	Jump Clustering	Intraday Periodicity
SV					x
ASV	x				x
SVJ	x	x			x
SVIJ	x	x	x		x
SVIJ-MHP	x	x	x	x	x

Table 3.4: Parameters Estimated in Stochastic Volatility Models

	AAPL									
	SV		ASV		SVJ		SVIJ		SVIJ-MHP	
	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.
$\mu(*10^{-5})$	10.35	5.714	7.112	6.214	6.761	5.749	6.257	5.861	5.990	5.591
κ	0.036	0.001	0.036	0.001	0.041	0.003	0.041	0.003	0.041	0.003
$\theta(*10^{-5})$	79.78	1.223	79.93	1.347	63.34	1.890	61.73	2.475	61.20	2.338
$\sigma_v(*10^{-3})$	6.357	0.786	6.367	0.913	5.670	1.088	5.659	0.885	5.614	0.807
ρ			-0.190	0.049	-0.183	0.049	-0.180	0.048	-0.184	0.049
μ_{Y+}					0.035	0.007	0.034	0.007	0.034	0.008
σ_{Y+}					0.045	0.006	0.045	0.005	0.045	0.005
μ_{Y-}					-0.037	0.009	-0.038	0.009	-0.037	0.009
σ_{Y-}					0.046	0.006	0.047	0.006	0.046	0.005
μ_V							0.561	0.063	0.560	0.063
DIC	-384719		-384694		-449190		-450340		-457330	
logL	201508		201444		216806		217280		221039	
	SPX									
	SV		ASV		SVJ		SVIJ		SVIJ-MHP	
	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.
$\mu(*10^{-5})$	12.91	1.552	6.804	1.589	5.346	1.228	5.352	1.207	5.737	1.191
κ	0.112	0.002	0.111	0.002	0.119	0.002	0.121	0.003	0.123	0.002
$\theta(*10^{-5})$	4.596	0.048	4.628	0.050	4.307	0.046	4.240	0.060	4.241	0.060
$\sigma_v(*10^{-3})$	3.025	0.305	3.025	0.310	2.969	0.280	2.965	0.302	2.973	0.292
ρ			-0.166	0.049	-0.163	0.049	-0.164	0.049	-0.162	0.048
μ_{Y+}					0.010	0.002	0.010	0.002	0.010	0.002
σ_{Y+}					0.015	0.002	0.015	0.002	0.015	0.002
μ_{Y-}					-0.012	0.003	-0.012	0.003	-0.012	0.003
σ_{Y-}					0.013	0.002	0.013	0.002	0.013	0.001
μ_V							1.047	0.162	1.046	0.161
DIC	-634698		-634617		-648981		-649500		-650586	
logL	318536		318451		321166		321412		322391	

Notes: DIC denotes deviance information criterion and LogL denotes log-likelihood. Parameter are estimated using five minutes returns data in the in-sample period (2012-2015) scaled by $M * 100$ ($M = 78$ for 5-minutes data).

Table 3.5: Parameters Estimated in the Marked Hawkes Process of SVIJ-MHP Model

	AAPL		SPX			AAPL		SPX	
	Post. mean	Post. std. dev.	Post. mean	Post. std. dev.		Post. mean	Post. std. dev.	Post. mean	Post. std. dev.
λ_0^{Y+}	**0.0088**	0.0003	**0.0018**	0.0001	*(Parameters in the impact function Eq. 3.4)*				
λ_0^{Y-}	**0.0101**	0.0004	**0.0026**	0.0002					
λ_0^{V}	**0.0011**	0.0001	**0.0006**	0.0001					
$\vartheta_{Y+,Y+}$	**0.1288**	0.0348	**0.1052**	0.0310	$\tilde{\alpha}_{Y+,Y+}$	**0.8168**	0.1983	**0.9679**	0.2827
$\vartheta_{Y-,Y+}$	**0.1855**	0.0449	**0.0956**	0.0336	$\tilde{\alpha}_{Y-,Y+}$	**0.7668**	0.1676	**0.9328**	0.2824
$\vartheta_{V,Y+}$	**0.4019**	0.1559	**0.2490**	0.0888	$\tilde{\alpha}_{V,Y+}$	0.1789	0.1598	0.2087	0.1546
$\vartheta_{Y+,Y-}$	**0.1005**	0.0325	**0.0510**	0.0077	$\tilde{\alpha}_{Y+,Y-}$	**0.8824**	0.2552	0.9255	0.6632
$\vartheta_{Y-,Y-}$	**0.1897**	0.0304	**0.2101**	0.0395	$\tilde{\alpha}_{Y-,Y-}$	**0.7244**	0.1323	**0.9173**	0.1706
$\vartheta_{V,Y-}$	0.3399	0.1862	0.2485	0.1509	$\tilde{\alpha}_{V,Y-}$	0.2543	0.2231	0.3880	0.2745
$\vartheta_{Y+,V}$	0.0048	0.0036	0.0247	0.0126	$\tilde{\alpha}_{Y+,v}$	0.7827	0.5705	0.9039	0.4674
$\vartheta_{Y-,V}$	**0.0118**	0.0060	**0.0174**	0.0050	$\tilde{\alpha}_{Y-,V}$	0.4633	0.3068	0.8864	0.6242
$\vartheta_{V,V}$	0.0951	0.0565	0.1211	0.0622	$\tilde{\alpha}_{V,V}$	0.3171	0.2121	0.3014	0.2030
$\beta_{Y+,Y+}$	**0.0947**	0.0136	**0.1730**	0.0427	$\tilde{\beta}_{Y+,Y+}$	**164.88**	64.589	**97.591**	35.554
$\beta_{Y-,Y+}$	**0.0786**	0.0094	**0.0694**	0.0168	$\tilde{\beta}_{Y-,Y+}$	**192.54**	61.447	174.18	178.24
$\beta_{V,Y+}$	**0.1694**	0.0343	**0.2465**	0.0573	$\tilde{\beta}_{V,Y+}$	**26.065**	7.2412	**34.902**	8.9053
$\beta_{Y+,Y-}$	**0.0559**	0.0121	**0.2852**	0.1398	$\tilde{\beta}_{Y+,Y-}$	105.90	61.360	**226.41**	114.14
$\beta_{Y-,Y-}$	**0.0955**	0.0107	**0.0890**	0.0144	$\tilde{\beta}_{Y-,Y-}$	**227.55**	23.3034	**214.19**	44.717
$\beta_{V,Y-}$	**0.1188**	0.0380	**0.2010**	0.0639	$\tilde{\beta}_{V,Y-}$	**23.673**	10.307	26.995	14.668
$\beta_{Y+,V}$	**0.3086**	0.0997	**0.3787**	0.1284	$\tilde{\beta}_{Y+,V}$	195.59	164.16	**292.20**	138.67
$\beta_{Y-,V}$	**0.3097**	0.0822	**0.5744**	0.2220	$\tilde{\beta}_{Y-,V}$	**443.14**	164.66	**294.28**	134.65
$\beta_{V,V}$	**0.3324**	0.0895	**0.3435**	0.1196	$\tilde{\beta}_{V,V}$	21.678	12.127	**30.812**	13.691

Notes: this table presents posterior means and standard deviations of parameters of the MHP kernel estimated in the in-sample period. Significant (95% of posterior is greater than 0) parameters are made bold.

Table 3.6: Variance Forecasts Performance of AAPL and SPX

AAPL										
	MSE					BIAS				
τ	SVIJ-MHP	SVIJ	SVJ	ASV	SV	SVIJ-MHP	SVIJ	SVJ	ASV	SV
1	**0.978**	0.991	0.998	0.997	1	**0.000**	-0.001	-0.002	-0.002	-0.002
6	**0.940**	0.984	1.004	1.005	1	**-0.005**	-0.006	-0.012	-0.012	-0.012
12	**0.936**	0.976	0.998	1.000	1	**-0.005**	-0.005	-0.008	-0.008	-0.008
24	**0.931**	0.972	1.003	1.004	1	-0.005	**-0.005**	-0.008	-0.007	-0.008
36	**0.926**	0.967	0.996	0.999	1	**-0.005**	-0.005	-0.007	-0.007	-0.007
78	**0.932**	0.961	0.999	1.001	1	**-0.006**	-0.007	-0.009	-0.009	-0.009
	R^2_{mz}					DM test				
τ	SVIJ-MHP	SVIJ	SVJ	ASV	SV		SVIJ	SVJ	ASV	SV
1	99.3%	**99.3%**	99.3%	99.3%	99.3%		0.0000	0.0000	0.0000	0.0000
6	**97.8%**	97.6%	97.6%	97.6%	97.6%		0.0000	0.0000	0.0000	0.0000
12	**95.0%**	94.8%	94.8%	94.7%	94.7%		0.0000	0.0000	0.0000	0.0000
24	**90.8%**	90.1%	90.1%	90.0%	90.2%		0.0000	0.0000	0.0000	0.0000
36	**88.0%**	87.0%	87.0%	87.1%	87.1%		0.0001	0.0000	0.0000	0.0000
78	**83.7%**	81.2%	81.0%	81.0%	81.2%		0.0000	0.0000	0.0000	0.0000
SPX										
	MSE					BIAS				
τ	SVIJ-MHP	SVIJ	SVJ	ASV	SV	SVIJ-MHP	SVIJ	SVJ	ASV	SV
1	**0.978**	0.996	0.999	1.001	1	**-0.001**	-0.002	-0.004	-0.004	-0.004
6	**0.966**	0.995	0.999	1.000	1	**-0.005**	-0.006	-0.007	-0.007	-0.007
12	**0.966**	0.991	0.999	1.000	1	-0.006	**-0.006**	-0.007	-0.007	-0.007
24	**0.965**	0.991	1.000	1.001	1	**-0.004**	-0.004	-0.005	-0.005	-0.005
36	**0.967**	0.990	1.000	0.999	1	**-0.006**	-0.007	-0.007	-0.007	-0.007
78	**0.969**	0.990	1.001	1.001	1	**-0.008**	-0.008	-0.009	-0.009	-0.009
	R^2_{mz}					DM test				
τ	SVIJ-MHP	SVIJ	SVJ	ASV	SV		SVIJ	SVJ	ASV	SV
1	**96.4%**	96.4%	96.4%	96.4%	96.4%		0.0022	0.0000	0.0002	0.0025
6	**84.8%**	84.5%	84.5%	84.5%	84.4%		0.0034	0.0002	0.0003	0.0005
12	**82.7%**	82.3%	82.3%	82.3%	82.2%		0.0009	0.0000	0.0000	0.0000
24	**75.2%**	74.7%	74.7%	74.7%	74.4%		0.0007	0.0000	0.0000	0.0000
36	**67.5%**	67.0%	66.8%	66.7%	67.0%		0.0025	0.0000	0.0001	0.0000
78	**62.8%**	61.1%	61.4%	61.1%	61.1%		0.0024	0.0001	0.0001	0.0002

Notes: R^2_{mz} denotes R^2 of Mincer-Zarnowitz regressions, DM test reports p-value of DM test of forecasts by SVIJ-MHP model against those by other models. MSE of forecasts by SV model are scaled to 1. Best performing models are marked bold.

Table 3.7: Conditional Performance of Variance Forecasts Using AAPL 5-minutes Data

After Positive Return Jumps

τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$	$h=78$	τ	Models	$h=1$	$h=2$	$h=3$
	SVIJ-MHP	0.942	0.941	0.950	0.949	0.946	0.949		SVIJ-MHP	0.935	0.937	0.935
	SVIJ	0.991	0.993	0.994	0.994	0.994	0.991		SVIJ	0.975	0.974	0.973
1	SVJ	0.989	0.991	0.993	0.992	0.994	0.996	24	SVJ	1.005	1.003	1.001
	ASV	0.984	0.989	0.988	0.991	0.993	0.995		ASV	1.010	1.006	1.004
	SV	1	1	1	1	1	1		SV	1	1	1
	SVIJ-MHP	0.923	0.931	0.936	0.932	0.939			SVIJ-MHP	0.935	0.938	
	SVIJ	0.985	0.990	0.988	0.986	0.987			SVIJ	0.970	0.972	
6	SVJ	0.999	1.002	1.003	1.002	1.001		36	SVJ	0.994	0.997	
	ASV	1.004	1.005	1.006	1.005	1.003			ASV	1.001	1.001	
	SV	1	1	1	1	1			SV	1	1	
	SVIJ-MHP	0.918	0.928	0.927	0.940				SVIJ-MHP	0.937		
	SVIJ	0.978	0.976	0.975	0.978				SVIJ	0.960		
12	SVJ	0.992	0.992	0.993	0.999			78	SVJ	0.998		
	ASV	0.998	0.996	0.995	1.000				ASV	1.001		
	SV	1	1	1	1				SV	1		

After Negative Return Jumps

τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$	$h=78$	τ	Models	$h=1$	$h=2$	$h=3$
	SVIJ-MHP	0.892	0.890	0.894	0.900	0.900	0.900		SVIJ-MHP	0.898	0.907	0.909
	SVIJ	0.999	1.001	0.999	0.998	0.998	0.993		SVIJ	0.975	0.975	0.976
1	SVJ	1.001	1.002	1.001	1.000	0.999	0.998	24	SVJ	1.006	1.007	1.005
	ASV	1.005	1.007	1.004	1.004	1.001	0.998		ASV	1.011	1.011	1.010
	SV	1	1	1	1	1	1		SV	1	1	1
	SVIJ-MHP	0.887	0.895	0.892	0.904	0.911			SVIJ-MHP	0.923	0.928	
	SVIJ	0.993	0.994	0.992	0.990	0.987			SVIJ	0.964	0.967	
6	SVJ	1.009	1.009	1.008	1.004	1.004		36	SVJ	0.991	0.995	
	ASV	1.006	1.007	1.008	1.007	1.006			ASV	0.994	0.998	
	SV	1	1	1	1	1			SV	1	1	
	SVIJ-MHP	0.885	0.891	0.899	0.906				SVIJ-MHP	0.937		
	SVIJ	0.984	0.980	0.980	0.979				SVIJ	0.961		
12	SVJ	0.994	0.996	0.998	0.998			78	SVJ	0.998		
	ASV	0.997	0.998	1.000	0.999				ASV	1.001		
	SV	1	1	1	1				SV	1		

After Variance Jumps

τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$	$h=78$	τ	Models	$h=1$	$h=2$	$h=3$
	SVIJ-MHP	0.947	0.951	0.976	0.981	0.979	0.976		SVIJ-MHP	0.883	0.924	0.928
	SVIJ	1.002	1.001	0.998	1.000	1.000	0.996		SVIJ	0.986	0.977	0.980
1	SVJ	0.999	0.999	1.000	1.001	1.001	0.998	24	SVJ	1.038	1.020	1.010
	ASV	1.008	1.010	1.006	1.006	1.005	1.001		ASV	1.040	1.027	1.017
	SV	1	1	1	1	1	1		SV	1	1	1
	SVIJ-MHP	0.877	0.894	0.894	0.907	0.925			SVIJ-MHP	0.928	0.939	
	SVIJ	1.004	1.006	1.003	1.001	0.996			SVIJ	0.972	0.975	
6	SVJ	1.010	1.009	1.006	1.004	0.999		36	SVJ	1.000	0.996	
	ASV	1.010	1.013	1.010	1.011	1.008			ASV	1.013	1.007	
	SV	1	1	1	1	1			SV	1	1	
	SVIJ-MHP	0.877	0.882	0.891	0.912				SVIJ-MHP	0.946		
	SVIJ	0.991	0.993	0.993	0.994				SVIJ	0.961		
12	SVJ	0.988	0.993	0.998	1.001			78	SVJ	0.994		
	ASV	0.990	0.993	0.995	0.999				ASV	0.995		
	SV	1	1	1	1				SV	1		

Notes: The table presents variance forecast performance by extracting h forecasts after different types of jumps. The purpose is to evaluate the forecasting contribution of jump clustering effect produced by different types of jumps.

Table 3.8: p-value of DM test on Variance Forecasts by SVIJ-MHP Model against Others (AAPL 5-min)

After Positive Return Jumps

τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$	$h=78$	τ	Models	$h=1$	$h=2$	$h=3$
1	SVIJ	0.071	0.011	0.003	0.000	0.002	0.000	24	SVIJ	0.000	0.000	0.000
	SVJ	0.063	0.008	0.001	0.000	0.000	0.000		SVJ	0.000	0.000	0.000
	ASV	0.027	0.001	0.000	0.000	0.000	0.000		ASV	0.000	0.000	0.000
	SV	0.051	0.001	0.000	0.000	0.000	0.000		SV	0.000	0.000	0.000
6	SVIJ	0.057	0.056	0.040	0.012	0.007		36	SVIJ	0.001	0.000	
	SVJ	0.037	0.035	0.021	0.003	0.001			SVJ	0.000	0.000	
	ASV	0.029	0.028	0.015	0.002	0.001			ASV	0.000	0.000	
	SV	0.037	0.038	0.022	0.003	0.001			SV	0.000	0.000	
12	SVIJ	0.001	0.000	0.000	0.000			78	SVIJ	0.000		
	SVJ	0.000	0.000	0.000	0.000				SVJ	0.000		
	ASV	0.000	0.000	0.000	0.000				ASV	0.000		
	SV	0.000	0.000	0.000	0.000				SV	0.000		

After Negative Return Jumps

τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$	$h=78$	τ	Models	$h=1$	$h=2$	$h=3$
1	SVIJ	0.006	0.005	0.006	0.004	0.000	0.000	24	SVIJ	0.000	0.000	0.000
	SVJ	0.006	0.005	0.006	0.004	0.000	0.000		SVJ	0.000	0.000	0.000
	ASV	0.004	0.003	0.003	0.002	0.000	0.000		ASV	0.000	0.000	0.000
	SV	0.008	0.006	0.007	0.004	0.000	0.000		SV	0.000	0.000	0.000
6	SVIJ	0.002	0.001	0.001	0.000	0.000		36	SVIJ	0.002	0.000	
	SVJ	0.001	0.000	0.000	0.000	0.000			SVJ	0.000	0.000	
	ASV	0.001	0.000	0.000	0.000	0.000			ASV	0.000	0.000	
	SV	0.001	0.000	0.000	0.000	0.000			SV	0.000	0.000	
12	SVIJ	0.001	0.001	0.000	0.000			78	SVIJ	0.000		
	SVJ	0.000	0.000	0.000	0.000				SVJ	0.000		
	ASV	0.000	0.000	0.000	0.000				ASV	0.000		
	SV	0.000	0.000	0.000	0.000				SV	0.000		

After Variance Jumps

τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$	$h=78$	τ	Models	$h=1$	$h=2$	$h=3$
1	SVIJ	0.039	0.056	0.027	0.023	0.004	0.000	24	SVIJ	0.037	0.017	0.003
	SVJ	0.041	0.059	0.027	0.024	0.004	0.000		SVJ	0.019	0.006	0.001
	ASV	0.030	0.041	0.021	0.019	0.002	0.000		ASV	0.020	0.003	0.001
	SV	0.043	0.066	0.030	0.032	0.006	0.000		SV	0.017	0.004	0.001
6	SVIJ	0.020	0.020	0.010	0.005	0.001		36	SVIJ	0.050	0.048	
	SVJ	0.020	0.018	0.008	0.004	0.000			SVJ	0.026	0.010	
	ASV	0.020	0.015	0.006	0.003	0.000			ASV	0.027	0.014	
	SV	0.019	0.015	0.007	0.003	0.000			SV	0.028	0.016	
12	SVIJ	0.023	0.019	0.014	0.027			78	SVIJ	0.041		
	SVJ	0.019	0.011	0.007	0.016				SVJ	0.021		
	ASV	0.019	0.012	0.007	0.014				ASV	0.023		
	SV	0.019	0.012	0.007	0.012				SV	0.025		

<0.01	[0.01,0.05)	[0.05, 0.1)	>0.1

Notes: Values in different intervals are colored differently. Results are under AAPL 5-min data. Denoting the difference of squared forecast error between two models as $d_t = e_{t,1}^2 - e_{t,2}^2$, the null hypothesis of the test is $E(d_t) = 0, \forall t$. A smaller p-value indicates a higher likelihood rejecting the null.

Table 3.9: Conditional Performance of Variance Forecasts Using SPX 5-minutes Data

After Positive Return Jumps

τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$	$h=78$	τ	Models	$h=1$	$h=2$	$h=3$
	SVIJ-MHP	0.972	0.965	0.961	0.994	0.995	0.995		SVIJ-MHP	0.979	0.981	0.981
	SVIJ	0.995	0.991	0.990	0.998	0.999	0.998		SVIJ	0.997	0.992	0.991
1	SVJ	0.991	1.000	0.998	0.999	0.999	0.999	24	SVJ	1.006	1.001	0.999
	ASV	1.002	1.005	1.004	1.000	1.000	1.001		ASV	1.009	1.003	1.000
	SV	1	1	1	1	1	1		SV	1	1	1
	SVIJ-MHP	0.933	0.938	0.944	0.946	0.949			SVIJ-MHP	0.979	0.985	
	SVIJ	0.982	0.981	0.981	0.981	0.979			SVIJ	0.990	0.991	
6	SVJ	0.998	0.996	0.997	0.996	0.998		36	SVJ	0.994	0.998	
	ASV	1.007	1.004	1.004	1.003	1.002			ASV	0.995	0.998	
	SV	1	1	1	1	1			SV	1	1	
	SVIJ-MHP	0.959	0.957	0.959	0.961				SVIJ-MHP	0.988		
	SVIJ	0.997	0.995	0.992	0.994				SVIJ	0.990		
12	SVJ	1.000	0.998	0.997	0.997			78	SVJ	1.003		
	ASV	1.002	0.998	0.998	0.998				ASV	1.001		
	SV	1	1	1	1				SV	1		

After Negative Return Jumps

τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$	$h=78$	τ	Models	$h=1$	$h=2$	$h=3$
	SVIJ-MHP	0.924	0.935	0.944	0.954	0.954	0.955		SVIJ-MHP	0.982	0.984	0.981
	SVIJ	0.977	0.977	0.977	0.976	0.977	0.977		SVIJ	0.996	0.998	0.994
1	SVJ	0.997	0.997	0.997	0.997	0.998	0.998	24	SVJ	1.005	1.007	1.003
	ASV	1.001	1.001	1.002	1.002	1.002	1.002		ASV	1.006	1.009	1.005
	SV	1	1	1	1	1	1		SV	1	1	1
	SVIJ-MHP	0.986	0.985	0.984	0.985	0.986			SVIJ-MHP	0.980	0.983	
	SVIJ	1.000	1.000	0.999	0.999	0.998			SVIJ	0.991	0.992	
6	SVJ	0.999	0.998	0.998	0.998	0.999		36	SVJ	0.996	0.998	
	ASV	1.000	1.001	1.001	1.001	1.002			ASV	0.997	0.999	
	SV	1	1	1	1	1			SV	1	1	
	SVIJ-MHP	0.975	0.980	0.980	0.980				SVIJ-MHP	0.987		
	SVIJ	0.996	0.996	0.995	0.994				SVIJ	0.993		
12	SVJ	0.997	0.999	0.999	0.999			78	SVJ	1.002		
	ASV	0.999	1.000	1.002	1.001				ASV	1.002		
	SV	1	1	1	1				SV	1		

After Variance Jumps

τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$	$h=78$	τ	Models	$h=1$	$h=2$	$h=3$
	SVIJ-MHP	0.961	0.958	0.957	0.956	0.960	0.961		SVIJ-MHP	0.976	0.982	0.981
	SVIJ	0.995	0.995	0.995	0.993	0.996	0.997		SVIJ	1.012	1.001	0.996
1	SVJ	0.994	0.995	0.996	0.996	0.998	0.998	24	SVJ	1.026	1.017	1.006
	ASV	1.001	1.001	1.002	1.002	1.002	1.002		ASV	1.034	1.022	1.011
	SV	1	1	1	1	1	1		SV	1	1	1
	SVIJ-MHP	0.956	0.952	0.953	0.954	0.958			SVIJ-MHP	0.967	0.978	
	SVIJ	1.004	1.003	1.004	1.002	1.001			SVIJ	0.992	0.995	
6	SVJ	0.999	0.997	0.998	0.998	1.001		36	SVJ	0.991	0.997	
	ASV	1.006	1.004	1.005	1.007	1.007			ASV	0.988	0.994	
	SV	1	1	1	1	1			SV	1	1	
	SVIJ-MHP	0.977	0.975	0.977	0.979				SVIJ-MHP	0.987		
	SVIJ	0.999	0.998	0.998	0.999				SVIJ	0.995		
12	SVJ	1.001	1.000	1.001	1.001			78	SVJ	1.007		
	ASV	0.999	0.998	0.999	0.998				ASV	1.003		
	SV	1	1	1	1				SV	1		

Notes: The table presents variance forecast performance by extracting h forecasts after different types of jumps. The purpose is to evaluate the forecasting contribution of jump clustering effect produced by different types of jumps.

Table 3.10: p-value of DM test on Variance Forecasts by SVIJ-MHP Model against Others (SPX 5-min)

After Positive Return Jumps

τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$	$h=78$	τ	Models	$h=1$	$h=2$	$h=3$
1	SVIJ	0.073	0.027	0.021	0.008	0.017	0.009	24	SVIJ	0.087	0.067	0.007
	SVJ	0.075	0.053	0.045	0.024	0.021	0.004		SVJ	0.051	0.029	0.004
	ASV	0.063	0.048	0.037	0.041	0.015	0.001		ASV	0.060	0.041	0.009
	SV	0.061	0.022	0.012	0.010	0.003	0.000		SV	0.050	0.026	0.004
6	SVIJ	0.085	0.078	0.085	0.065	0.009		36	SVIJ	0.089	0.076	
	SVJ	0.103	0.095	0.088	0.065	0.003			SVJ	0.013	0.002	
	ASV	0.071	0.069	0.065	0.043	0.003			ASV	0.012	0.002	
	SV	0.081	0.074	0.074	0.048	0.003			SV	0.035	0.012	
12	SVIJ	0.068	0.041	0.044	0.015			78	SVIJ	0.082		
	SVJ	0.055	0.037	0.020	0.011				SVJ	0.004		
	ASV	0.032	0.022	0.008	0.002				ASV	0.004		
	SV	0.053	0.031	0.015	0.004				SV	0.003		

After Negative Return Jumps

τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$	$h=78$	τ	Models	$h=1$	$h=2$	$h=3$
1	SVIJ	0.021	0.012	0.011	0.008	0.004	0.001	24	SVIJ	0.050	0.048	0.040
	SVJ	0.014	0.009	0.009	0.008	0.004	0.000		SVJ	0.062	0.049	0.022
	ASV	0.048	0.069	0.085	0.065	0.042	0.003		ASV	0.064	0.051	0.026
	SV	0.026	0.038	0.055	0.038	0.030	0.002		SV	0.052	0.027	0.013
6	SVIJ	0.028	0.026	0.026	0.035	0.017		36	SVIJ	0.070	0.028	
	SVJ	0.026	0.021	0.020	0.024	0.006			SVJ	0.014	0.011	
	ASV	0.025	0.023	0.017	0.022	0.006			ASV	0.007	0.012	
	SV	0.030	0.025	0.017	0.024	0.009			SV	0.032	0.013	
12	SVIJ	0.052	0.023	0.016	0.010			78	SVIJ	0.048		
	SVJ	0.046	0.021	0.018	0.007				SVJ	0.021		
	ASV	0.015	0.023	0.020	0.004				ASV	0.021		
	SV	0.032	0.017	0.009	0.002				SV	0.016		

After Variance Jumps

τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$	$h=78$	τ	Models	$h=1$	$h=2$	$h=3$
1	SVIJ	0.062	0.029	0.026	0.021	0.007	0.005	24	SVIJ	0.170	0.173	0.096
	SVJ	0.083	0.028	0.024	0.015	0.006	0.005		SVJ	0.107	0.086	0.052
	ASV	0.072	0.054	0.054	0.041	0.023	0.018		ASV	0.124	0.118	0.078
	SV	0.060	0.035	0.034	0.027	0.014	0.012		SV	0.091	0.053	0.019
6	SVIJ	0.066	0.071	0.069	0.067	0.023		36	SVIJ	0.135	0.065	
	SVJ	0.100	0.090	0.082	0.068	0.020			SVJ	0.073	0.023	
	ASV	0.081	0.078	0.075	0.061	0.022			ASV	0.027	0.027	
	SV	0.089	0.088	0.087	0.077	0.027			SV	0.112	0.060	
12	SVIJ	0.088	0.060	0.063	0.049			78	SVIJ	0.078		
	SVJ	0.083	0.063	0.063	0.041				SVJ	0.042		
	ASV	0.053	0.043	0.041	0.026				ASV	0.038		
	SV	0.084	0.056	0.057	0.032				SV	0.052		

<0.01	[0.01,0.05)	[0.05, 0.1)	>0.1

Notes: Values in different intervals are colored differently. Results are under SPX 5-min data. Denoting the difference of squared forecast error between two models as $d_t = e_{t,1}^2 - e_{t,2}^2$, the null hypothesis of the test is $E(d_t) = 0, \forall t$. A smaller p-value indicates a higher likelihood rejecting the null.

Table 3.11: p-value of DM test on Variance Forecasts by SVIJ-MHP Model against Others (AAPL,SPX 10-min)

After Positive Return Jumps

	AAPL						SPX					
τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$	τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$
1	SVIJ	0.051	0.062	0.045	0.035	0.058	1	SVIJ	0.0863	0.1063	0.0993	0.0967	0.0725
	SVJ	0.052	0.055	0.078	0.040	0.011		SVJ	0.0752	0.1026	0.0767	0.0609	0.0723
	ASV	0.036	0.019	0.019	0.003	0.001		ASV	0.0520	0.0949	0.0980	0.1132	0.1513
	SV	0.017	0.018	0.014	0.002	0.001		SV	0.0784	0.0915	0.1052	0.0890	0.1110
6	SVIJ	0.008	0.002	0.001	0.005	0.000	6	SVIJ	0.4747	0.7910	0.1621	0.2161	0.1559
	SVJ	0.006	0.000	0.000	0.000	0.000		SVJ	0.2944	0.9349	0.2374	0.1055	0.0184
	ASV	0.008	0.002	0.001	0.000	0.000		ASV	0.1899	0.3206	0.0947	0.1045	0.0660
	SV	0.003	0.001	0.000	0.000	0.000		SV	0.1262	0.0897	0.1100	0.0830	0.0483
12	SVIJ	0.075	0.088	0.036			12	SVIJ	0.8547	0.8330	0.6956		
	SVJ	0.005	0.000	0.000				SVJ	0.1177	0.0709	0.0579		
	ASV	0.007	0.002	0.002				ASV	0.1112	0.0791	0.0456		
	SV	0.006	0.000	0.001				SV	0.6711	0.1589	0.1063		
39	SVIJ	0.016					39	SVIJ	0.7143				
	SVJ	0.004						SVJ	0.1051				
	ASV	0.005						ASV	0.7321				
	SV	0.005						SV	0.7684				

After Negative Return Jumps

	AAPL						SPX					
τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$	τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$
1	SVIJ	0.015	0.007	0.007	0.005	0.001	1	SVIJ	0.066	0.061	0.061	0.057	0.066
	SVJ	0.008	0.003	0.002	0.001	0.000		SVJ	0.064	0.060	0.059	0.054	0.064
	ASV	0.012	0.006	0.005	0.003	0.000		ASV	0.066	0.061	0.061	0.057	0.067
	SV	0.010	0.004	0.002	0.002	0.000		SV	0.066	0.061	0.061	0.056	0.065
6	SVIJ	0.020	0.010	0.006	0.002	0.000	6	SVIJ	0.055	0.050	0.061	0.049	0.034
	SVJ	0.014	0.003	0.001	0.000	0.000		SVJ	0.041	0.055	0.056	0.029	0.015
	ASV	0.009	0.002	0.001	0.000	0.000		ASV	0.026	0.027	0.041	0.026	0.008
	SV	0.013	0.003	0.001	0.000	0.000		SV	0.030	0.042	0.047	0.033	0.009
12	SVIJ	0.036	0.012	0.010			12	SVIJ	0.607	0.697	0.160		
	SVJ	0.015	0.001	0.000				SVJ	0.090	0.111	0.048		
	ASV	0.017	0.002	0.001				ASV	0.062	0.066	0.026		
	SV	0.015	0.002	0.000				SV	0.018	0.050	0.016		
39	SVIJ	0.015					39	SVIJ	0.796				
	SVJ	0.002						SVJ	0.086				
	ASV	0.001						ASV	0.182				
	SV	0.001						SV	0.454				

<0.01	[0.01,0.05)	[0.05, 0.1)	>0.1

Notes: Values in different intervals are colored differently. Results are under AAPL and SPX 10-min data.

Table 3.12: p-value of DM test on Variance Forecasts by SVIJ-MHP Model against Others (AAPL,SPX 30-min)

After Positive Return Jumps

AAPL							SPX						
τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$	τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$
1	SVIJ	0.191	0.189	0.146	0.130	0.082	1	SVIJ	0.0109	0.4123	0.1182	0.3637	0.4136
	SVJ	0.077	0.029	0.017	0.006	0.001		SVJ	0.0133	0.0931	0.4192	0.6061	0.2350
	ASV	0.054	0.051	0.022	0.002	0.000		ASV	0.0027	0.0138	0.0569	0.1051	0.1528
	SV	0.079	0.043	0.026	0.003	0.001		SV	0.0093	0.0629	0.0757	0.0762	0.4093
2	SVIJ	0.572	0.844	0.400	0.683		2	SVIJ	0.1229	0.1653	0.2042	0.2193	
	SVJ	0.061	0.016	0.004	0.004			SVJ	0.2025	0.6089	0.4786	0.0905	
	ASV	0.096	0.053	0.019	0.002			ASV	0.1365	0.2130	0.1958	0.1497	
	SV	0.137	0.063	0.016	0.001			SV	0.5966	0.9127	0.8047	0.2922	
6	SVIJ	0.125	0.109				6	SVIJ	0.1308	0.0699			
	SVJ	0.140	0.100					SVJ	0.0349	0.1263			
	ASV	0.101	0.098					ASV	0.2532	0.1034			
	SV	0.099	0.111					SV	0.1290	0.0739			
13	SVIJ	0.119					13	SVIJ	0.2147				
	SVJ	0.111						SVJ	0.0855				
	ASV	0.111						ASV	0.0736				
	SV	0.115						SV	0.0878				

After Negative Return Jumps

AAPL							SPX						
τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$	τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$
1	SVIJ	0.070	0.064	0.034	0.139	0.034	1	SVIJ	0.032	0.107	0.117	0.121	0.082
	SVJ	0.075	0.059	0.013	0.011	0.009		SVJ	0.086	0.129	0.132	0.140	0.069
	ASV	0.036	0.034	0.010	0.010	0.005		ASV	0.061	0.095	0.097	0.122	0.054
	SV	0.052	0.048	0.013	0.026	0.017		SV	0.054	0.131	0.131	0.142	0.093
2	SVIJ	0.041	0.146	0.176	0.076		2	SVIJ	0.103	0.072	0.112	0.207	
	SVJ	0.028	0.043	0.034	0.003			SVJ	0.124	0.094	0.134	0.106	
	ASV	0.029	0.055	0.088	0.004			ASV	0.119	0.099	0.129	0.139	
	SV	0.003	0.075	0.052	0.004			SV	0.135	0.101	0.126	0.130	
6	SVIJ	0.108	0.011				6	SVIJ	0.105	0.106			
	SVJ	0.030	0.030					SVJ	0.120	0.113			
	ASV	0.006	0.003					ASV	0.239	0.203			
	SV	0.042	0.050					SV	0.219	0.105			
13	SVIJ	0.587					13	SVIJ	0.048				
	SVJ	0.018						SVJ	0.056				
	ASV	0.029						ASV	0.053				
	SV	0.054						SV	0.069				

<0.01	[0.01,0.05)	[0.05, 0.1)	>0.1

Notes: Values in different intervals are colored differently. Results are under AAPL and SPX 30-min data.

Table 3.13: p-value of DM test on Variance Forecasts by SVIJ-MHP Model against Others (after variance jumps)

10-minutes Data

		AAPL							SPX					
τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$	τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$
1	SVIJ	0.091	0.051	0.057	0.037	0.019	1	SVIJ	0.0874	0.0705	0.0415	0.0684	0.0211
	SVJ	0.080	0.048	0.052	0.034	0.012		SVJ	0.0826	0.0992	0.0889	0.0922	0.0117
	ASV	0.095	0.054	0.057	0.035	0.018		ASV	0.0465	0.0571	0.0415	0.0614	0.0102
	SV	0.091	0.056	0.063	0.039	0.024		SV	0.0481	0.0501	0.0605	0.0744	0.0080
6	SVIJ	0.089	0.080	0.081	0.093	0.096	6	SVIJ	0.1449	0.1453	0.1473	0.1382	0.1367
	SVJ	0.090	0.081	0.077	0.064	0.040		SVJ	0.1442	0.1446	0.1478	0.1375	0.1357
	ASV	0.086	0.057	0.052	0.042	0.017		ASV	0.1458	0.1460	0.1462	0.1384	0.1385
	SV	0.091	0.067	0.066	0.068	0.053		SV	0.1448	0.1449	0.1448	0.1372	0.1296
12	SVIJ	0.125	0.134	0.128			12	SVIJ	0.8143	0.7631	0.5032		
	SVJ	0.114	0.101	0.075				SVJ	0.1324	0.1442	0.0411		
	ASV	0.116	0.110	0.090				ASV	0.1020	0.0761	0.0190		
	SV	0.122	0.109	0.090				SV	0.6783	0.5579	0.1547		
39	SVIJ	0.473					39	SVIJ	0.8153				
	SVJ	0.060						SVJ	0.1107				
	ASV	0.063						ASV	0.4610				
	SV	0.029						SV	0.5353				

30-minutes Data

		AAPL							SPX					
τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$	τ	Models	$h=1$	$h=2$	$h=3$	$h=6$	$h=13$
1	SVIJ	0.562	0.420	0.359	0.317	0.516	1	SVIJ	0.172	0.084	0.100	0.122	0.222
	SVJ	0.090	0.045	0.070	0.040	0.005		SVJ	0.477	0.221	0.245	0.366	0.235
	ASV	0.069	0.001	0.073	0.096	0.023		ASV	0.196	0.116	0.131	0.191	0.117
	SV	0.066	0.031	0.131	0.121	0.036		SV	0.195	0.208	0.789	0.415	0.755
2	SVIJ	0.066	0.031	0.023	0.041		2	SVIJ	0.396	0.785	0.822	0.836	
	SVJ	0.076	0.094	0.040	0.100			SVJ	0.618	0.773	0.817	0.517	
	ASV	0.084	0.021	0.025	0.005			ASV	0.140	0.771	0.944	0.390	
	SV	0.109	0.078	0.061	0.133			SV	0.517	0.786	0.794	0.923	
6	SVIJ	0.123	0.207				6	SVIJ	0.134	0.125			
	SVJ	0.175	0.166					SVJ	0.140	0.130			
	ASV	0.104	0.077					ASV	0.139	0.192			
	SV	0.150	0.124					SV	0.142	0.237			
13	SVIJ	0.294					13	SVIJ	0.136				
	SVJ	0.277						SVJ	0.140				
	ASV	0.241						ASV	0.135				
	SV	0.276						SV	0.139				

<0.01	[0.01,0.05)	[0.05, 0.1)	>0.1

Notes: Values in different intervals are colored differently. Results are testing on variance forecasts after variance jumps.

Table 3.14: Daily Realised Volatility Forecasting Performance

	AAPL				SPX			
Model	R^2	MSE	BIAS	DM p-value	R^2	MSE	BIAS	DM p-value
SVIJ-MHP	**57.9%**	**0.94**	-0.010	N/A	**70.7%**	**0.97**	-0.011	N/A
SVIJ	50.9%	1.00	-0.013	0.012	69.6%	1.00	-0.011	0.006
SVJ	50.9%	1.00	-0.016	0.009	69.6%	1.00	-0.009	0.007
ASV	50.6%	1.00	-0.016	0.009	69.6%	1.00	-0.009	0.011
SV	50.9%	1.00	-0.016	0.012	69.6%	1.00	-0.009	0.006
SVIJ-MHP (b.e.c.)	54.3%	1.43	-0.067	0.000	66.3%	2.86	-0.031	0.000
SVIJ (b.e.c.)	48.0%	1.65	-0.073	0.000	62.5%	2.90	-0.033	0.000
SVJ (b.e.c.)	48.2%	1.70	-0.086	0.000	62.6%	2.93	-0.043	0.000
ASV (b.e.c.)	48.0%	1.71	-0.086	0.000	62.6%	2.93	-0.043	0.000
SV (b.e.c.)	48.1%	1.71	-0.086	0.000	62.8%	2.94	-0.043	0.000
HAR-RV	46.0%	1.08	-0.030	0.005	65.3%	1.14	-0.032	0.056
HAR-RV-CJ	46.3%	1.07	-0.032	0.005	65.3%	1.14	-0.031	0.066
HAR-RV-SJ	45.9%	1.08	-0.030	0.003	67.0%	1.08	-0.032	0.045
Realized-GARCH	45.4%	1.24	0.090	0.002	63.8%	1.40	0.083	0.001
Realized-GARCH-log	46.3%	1.18	0.086	0.003	65.6%	1.21	0.083	0.000
Realized-EGARCH	39.4%	1.33	0.088	0.022	69.0%	1.10	0.083	0.071
No-change Forecast	42.5%	1.29	**0.000**	0.009	63.5%	1.33	**0.001**	0.005

<0.01	[0.01,0.05)	[0.05, 0.1)	>0.1

Notes: The fifth column 'DM p-value' reports p-values of DM test on forecasts by SVIJ-MHP model against those by other models. 'b.e.c.' stands for models before error correction by the KF.

Table 3.15: Auxiliary Regression of Realised Volatility Forecast Errors

	AAPL				SPX			
Model	b_0		b_1		b_0		b_1	
SVIJ	0.01	(0.83)	-0.01	(-0.35)	0.00	(0.78)	0.00	(-0.17)
SVJ	0.01	(0.84)	-0.01	(-0.35)	0.00	(0.78)	0.00	(-0.17)
ASV	0.01	(0.88)	-0.01	(-0.31)	0.00	(0.78)	0.00	(-0.18)
SV	0.01	(0.85)	-0.01	(-0.35)	0.00	(0.78)	0.00	(-0.17)
SVIJ-MHP (b.e.c.)	0.00	(0.05)	0.03	(0.72)	0.07	(0.94)	-0.01	(-0.26)
SVIJ (b.e.c.)	0.00	(0.02)	0.03	(0.67)	0.01	(0.98)	-0.01	(-0.23)
SVJ (b.e.c.)	0.00	(0.06)	0.03	(0.71)	0.01	(0.99)	-0.01	(-0.24)
ASV (b.e.c.)	0.00	(0.04)	0.03	(0.69)	0.01	(0.96)	-0.01	(-0.21)
SV (b.e.c.)	0.00	(0.05)	0.03	(0.71)	0.01	(0.99)	-0.01	(-0.25)
HAR-RV	0.01	(0.89)	0.01	(0.16)	0.00	(1.29)	-0.01	(-0.49)
HAR-RV-CJ	0.01	(0.79)	0.01	(0.26)	0.00	(1.27)	-0.01	(-0.47)
HAR-RV-SJ	0.01	(0.82)	0.01	(0.24)	0.00	(0.48)	0.01	(0.46)
Realized-GARCH	0.01	(0.87)	-0.01	(-0.13)	0.00	(1.03)	-0.02	(-0.43)
Realized-GARCH-log	0.01	(0.92)	0.00	(-0.04)	0.00	(1.25)	-0.02	(-0.54)
Realized-EGARCH	0.01	(0.72)	0.01	(0.29)	0.00	(0.23)	0.02	(0.6)
No-change Forecast	0.01	(2.23)	0.00	(0.01)	0.00	(1.72)	-0.01	(-1.12)

Notes: Every two columns report OLS estimates and corresponding test-statistics (test on whether coefficient estimates equal to 0, 'b.e.c.' stands for models before error correction by the KF.)

Table 3.16: Expected Loss of VaR and ES and Correspond p-value of DM Tests (AAPL 5-min)

		Expected Loss					DM test			
τ	α	SVIJ-MHP	SVIJ	SVJ	ASV	SV	SVIJ	SVJ	ASV	SV
1	1.0%	0.900	0.979	0.990	0.990	1.000	0.000	0.000	0.000	0.000
	2.5%	0.924	0.986	1.001	1.000	1.000	0.000	0.000	0.000	0.000
	5.0%	0.935	0.990	1.004	1.003	1.000	0.000	0.000	0.000	0.000
	10.0%	0.941	0.990	1.005	1.004	1.000	0.000	0.000	0.000	0.000
6	1.0%	0.917	0.976	0.987	0.990	1.000	0.007	0.012	0.015	0.003
	2.5%	0.939	0.983	0.995	0.997	1.000	0.002	0.005	0.005	0.001
	5.0%	0.949	0.987	1.000	1.001	1.000	0.001	0.001	0.001	0.000
	10.0%	0.958	0.989	1.002	1.002	1.000	0.000	0.000	0.000	0.000
12	1.0%	0.885	0.962	0.978	0.974	1.000	0.044	0.019	0.038	0.010
	2.5%	0.922	0.977	0.991	0.992	1.000	0.023	0.011	0.015	0.003
	5.0%	0.938	0.983	0.996	0.996	1.000	0.010	0.004	0.006	0.001
	10.0%	0.947	0.986	1.000	0.999	1.000	0.003	0.001	0.001	0.000
24	1.0%	0.908	0.954	0.979	0.986	1.000	0.013	0.007	0.007	0.009
	2.5%	0.938	0.967	0.985	0.988	1.000	0.004	0.003	0.002	0.002
	5.0%	0.955	0.973	0.986	0.991	1.000	0.002	0.001	0.001	0.000
	10.0%	0.973	0.983	0.995	0.998	1.000	0.001	0.000	0.000	0.000
36	1.0%	0.878	0.956	0.956	0.969	1.000	0.009	0.013	0.005	0.003
	2.5%	0.922	0.972	0.975	0.980	1.000	0.004	0.003	0.002	0.001
	5.0%	0.946	0.986	0.983	0.986	1.000	0.001	0.001	0.001	0.000
	10.0%	0.967	0.995	0.992	0.994	1.000	0.000	0.000	0.000	0.000
78	1.0%	0.843	0.920	0.920	0.933	1.000	0.051	0.043	0.064	0.013
	2.5%	0.882	0.948	0.950	0.961	1.000	0.014	0.008	0.015	0.003
	5.0%	0.916	0.970	0.972	0.976	1.000	0.004	0.001	0.004	0.001
	10.0%	0.943	0.983	0.983	0.984	1.000	0.001	0.000	0.001	0.000

Notes: the left section of this table presents the joint expected losses of VaR and ES calculated by a loss function proposed by Fissler & Ziegel (2016). The right hand side reports the p-value of DM tests (Diebold & Mariano 1995) of the SVIJ-MHP model against other models. The data used is AAPL 5-minutes data.

Table 3.17: Coverage Ratio, DQ and DES Tests (AAPL 5-min)

τ	α_{ve}	SVIJ-MHP			SVIJ			SVJ			ASV			SV		
1	1.0%	0.8%	0.000	0.000	1.1%	0.000	0.000	1.1%	0.000	0.000	1.1%	0.000	0.000	1.2%	0.000	0.000
	2.5%	1.8%	0.000	0.000	2.1%	0.000	0.000	2.1%	0.000	0.000	2.1%	0.000	0.000	2.4%	0.000	0.000
	5.0%	3.8%	0.000	0.000	4.1%	0.000	0.000	4.1%	0.000	0.000	4.1%	0.000	0.000	4.7%	0.000	0.000
	10.0%	8.7%	0.000	0.000	8.8%	0.000	0.000	8.8%	0.000	0.000	8.9%	0.000	0.000	9.6%	0.000	0.000
6	1.0%	1.0%	0.003	0.002	1.2%	0.008	0.001	1.3%	0.001	0.000	1.2%	0.000	0.000	1.6%	0.000	0.000
	2.5%	2.2%	0.000	0.000	2.5%	0.000	0.000	2.4%	0.000	0.000	2.6%	0.000	0.000	3.1%	0.000	0.000
	5.0%	4.0%	0.000	0.000	4.4%	0.000	0.000	4.6%	0.000	0.000	4.6%	0.000	0.000	5.4%	0.000	0.000
	10.0%	7.9%	0.000	0.000	8.5%	0.000	0.000	8.5%	0.000	0.000	8.7%	0.000	0.000	9.9%	0.000	0.000
12	1.0%	1.2%	0.360	0.191	1.3%	0.061	0.032	1.3%	0.086	0.036	1.4%	0.053	0.016	1.9%	0.000	0.000
	2.5%	2.2%	0.003	0.026	2.5%	0.052	0.093	2.6%	0.028	0.059	2.6%	0.023	0.059	3.5%	0.000	0.000
	5.0%	4.1%	0.000	0.001	4.5%	0.000	0.014	4.6%	0.000	0.015	4.7%	0.000	0.016	5.6%	0.006	0.001
	10.0%	8.2%	0.000	0.000	8.8%	0.000	0.000	8.9%	0.000	0.000	9.2%	0.000	0.001	10.7%	0.007	0.001
24	1.0%	1.3%	0.000	0.000	1.4%	0.053	0.024	1.3%	0.170	0.047	1.6%	0.000	0.000	2.2%	0.000	0.000
	2.5%	2.4%	0.205	0.074	2.5%	0.014	0.013	2.8%	0.019	0.007	2.8%	0.009	0.003	3.8%	0.000	0.000
	5.0%	4.9%	0.433	0.181	5.1%	0.026	0.007	5.4%	0.022	0.004	5.6%	0.016	0.001	7.0%	0.000	0.000
	10.0%	8.5%	0.025	0.395	9.1%	0.002	0.009	9.4%	0.008	0.013	9.6%	0.018	0.002	12.2%	0.002	0.000
36	1.0%	1.3%	0.000	0.000	1.5%	0.000	0.000	1.8%	0.000	0.000	1.7%	0.000	0.000	2.4%	0.000	0.000
	2.5%	2.5%	0.615	0.461	3.0%	0.193	0.080	3.1%	0.150	0.057	3.1%	0.041	0.021	4.4%	0.001	0.000
	5.0%	5.0%	0.712	0.519	5.6%	0.232	0.062	5.9%	0.055	0.019	5.9%	0.058	0.010	7.8%	0.000	0.000
	10.0%	9.2%	0.555	0.741	9.6%	0.192	0.130	9.6%	0.401	0.178	10.1%	0.309	0.084	12.1%	0.014	0.000
78	1.0%	1.6%	0.029	0.023	1.8%	0.007	0.009	2.5%	0.001	0.001	2.5%	0.000	0.000	3.7%	0.000	0.000
	2.5%	3.0%	0.024	0.019	3.2%	0.003	0.003	3.5%	0.004	0.001	4.0%	0.007	0.001	4.7%	0.000	0.000
	5.0%	5.3%	0.204	0.081	5.7%	0.013	0.004	6.1%	0.003	0.001	6.1%	0.003	0.001	9.1%	0.000	0.000
	10.0%	10.0%	0.598	0.268	11.3%	0.000	0.000	11.4%	0.005	0.001	11.8%	0.001	0.000	14.2%	0.000	0.000

Notes:α_{ve} denotes significant levels of VaR and ES. Three columns of each model forecast results reports coverage ratio (left), p-value of DQ test (middle) and p-value of DES test respectively.

Table 3.18: Expected Loss of VaR and ES and Correspond p-value of DM Tests (SPX 5-min)

		Expected Loss					DM test			
τ	α	SVIJ-MHP	SVIJ	SVJ	ASV	SV	SVIJ	SVJ	ASV	SV
1	1.0%	0.926	1.082	1.082	1.083	1.000	0.002	0.015	0.001	0.001
	2.5%	0.851	1.010	1.010	1.001	1.000	0.000	0.001	0.000	0.000
	5.0%	0.832	0.948	0.936	0.843	1.000	0.000	0.000	0.000	0.000
	10.0%	0.857	0.982	0.987	0.994	1.000	0.000	0.000	0.000	0.000
6	1.0%	0.933	1.009	1.005	1.019	1.000	0.030	0.011	0.033	0.010
	2.5%	0.948	1.026	1.020	1.023	1.000	0.013	0.005	0.023	0.002
	5.0%	0.958	1.032	1.027	1.028	1.000	0.008	0.002	0.010	0.000
	10.0%	0.962	1.036	1.030	1.029	1.000	0.003	0.000	0.003	0.000
12	1.0%	0.907	0.990	0.985	0.996	1.000	0.004	0.002	0.000	0.000
	2.5%	0.921	0.994	0.990	0.997	1.000	0.001	0.001	0.000	0.000
	5.0%	0.932	1.004	1.000	1.004	1.000	0.000	0.000	0.000	0.000
	10.0%	0.939	1.008	1.004	1.007	1.000	0.000	0.000	0.000	0.000
24	1.0%	0.861	0.917	0.907	0.908	1.000	0.028	0.026	0.035	0.003
	2.5%	0.900	0.939	0.936	0.933	1.000	0.013	0.024	0.019	0.001
	5.0%	0.929	0.960	0.958	0.957	1.000	0.007	0.010	0.006	0.000
	10.0%	0.951	0.975	0.974	0.973	1.000	0.003	0.003	0.002	0.000
36	1.0%	0.815	0.867	0.871	0.868	1.000	0.008	0.018	0.007	0.001
	2.5%	0.876	0.901	0.909	0.911	1.000	0.012	0.005	0.003	0.000
	5.0%	0.916	0.938	0.941	0.940	1.000	0.011	0.004	0.002	0.000
	10.0%	0.943	0.962	0.963	0.962	1.000	0.005	0.002	0.001	0.000
78	1.0%	0.707	0.788	0.778	0.790	1.000	0.025	0.010	0.023	0.005
	2.5%	0.785	0.842	0.848	0.846	1.000	0.014	0.003	0.012	0.001
	5.0%	0.844	0.888	0.892	0.893	1.000	0.007	0.002	0.006	0.001
	10.0%	0.886	0.921	0.922	0.925	1.000	0.004	0.001	0.003	0.000

Notes: the left section of this table presents the joint expected losses of VaR and ES calculated by a loss function proposed by Fissler & Ziegel (2016). The right hand side reports the p-value of DM tests (Diebold & Mariano 1995) of the SVIJ-MHP model against other models. The data used is SPX 5-minutes data.

Table 3.19: Coverage Ratio, DQ and DES Tests (SPX 5-min)

τ	α_{ve}	SVIJ-MHP			SVIJ			SVJ			ASV			SV		
1	1.0%	0.3%	0.000	0.000	0.4%	0.000	0.000	0.4%	0.000	0.000	0.4%	0.000	0.000	0.5%	0.000	0.000
	2.5%	0.8%	0.000	0.000	0.9%	0.000	0.000	1.0%	0.000	0.000	0.9%	0.000	0.000	1.2%	0.000	0.000
	5.0%	2.1%	0.000	0.000	2.1%	0.000	0.000	2.2%	0.000	0.000	2.2%	0.000	0.000	2.6%	0.000	0.000
	10.0%	6.0%	0.000	0.000	6.1%	0.000	0.000	6.1%	0.000	0.000	6.2%	0.000	0.000	7.1%	0.000	0.000
6	1.0%	0.6%	0.000	0.000	0.7%	0.000	0.000	0.7%	0.000	0.000	0.7%	0.000	0.000	1.1%	0.000	0.000
	2.5%	1.6%	0.000	0.000	1.8%	0.000	0.000	1.8%	0.000	0.000	1.8%	0.000	0.000	2.5%	0.000	0.000
	5.0%	3.5%	0.000	0.000	3.7%	0.000	0.000	3.8%	0.000	0.000	3.8%	0.000	0.000	5.1%	0.000	0.000
	10.0%	7.8%	0.000	0.000	8.2%	0.000	0.000	8.2%	0.000	0.000	8.3%	0.000	0.000	10.2%	0.000	0.000
12	1.0%	1.0%	0.017	0.012	1.1%	0.000	0.000	1.1%	0.000	0.000	1.2%	0.000	0.000	1.6%	0.000	0.000
	2.5%	2.3%	0.003	0.001	2.5%	0.000	0.000	2.6%	0.000	0.000	2.6%	0.000	0.000	3.8%	0.000	0.000
	5.0%	4.5%	0.002	0.000	4.7%	0.000	0.000	4.8%	0.000	0.000	4.9%	0.000	0.000	6.3%	0.000	0.000
	10.0%	8.5%	0.000	0.000	8.8%	0.000	0.000	8.9%	0.000	0.000	9.2%	0.000	0.000	11.5%	0.000	0.000
24	1.0%	1.4%	0.114	0.097	1.6%	0.000	0.000	1.6%	0.000	0.000	1.6%	0.000	0.000	2.7%	0.000	0.000
	2.5%	2.8%	0.081	0.068	3.1%	0.000	0.000	3.4%	0.000	0.000	3.1%	0.000	0.000	4.6%	0.000	0.000
	5.0%	5.3%	0.020	0.025	5.4%	0.000	0.000	5.7%	0.000	0.000	5.8%	0.000	0.000	7.9%	0.000	0.000
	10.0%	9.5%	0.003	0.006	9.7%	0.000	0.000	9.6%	0.000	0.000	9.8%	0.000	0.000	13.2%	0.000	0.000
36	1.0%	1.7%	0.008	0.004	1.7%	0.001	0.000	1.8%	0.000	0.000	1.9%	0.000	0.000	3.5%	0.000	0.000
	2.5%	3.4%	0.006	0.002	3.7%	0.000	0.000	3.5%	0.000	0.000	4.1%	0.000	0.000	5.3%	0.000	0.000
	5.0%	6.0%	0.012	0.002	6.3%	0.000	0.000	6.3%	0.000	0.000	6.3%	0.000	0.000	8.6%	0.000	0.000
	10.0%	9.1%	0.010	0.006	9.5%	0.000	0.000	9.6%	0.001	0.000	9.9%	0.002	0.000	13.2%	0.000	0.000
78	1.0%	2.3%	0.000	0.000	2.6%	0.000	0.000	2.6%	0.000	0.000	2.7%	0.000	0.000	4.7%	0.000	0.000
	2.5%	4.0%	0.000	0.000	4.0%	0.000	0.000	4.1%	0.000	0.000	4.1%	0.000	0.000	6.3%	0.000	0.000
	5.0%	6.9%	0.004	0.000	7.2%	0.004	0.000	7.2%	0.004	0.000	7.5%	0.005	0.000	10.7%	0.000	0.000
	10.0%	9.9%	0.083	0.004	10.6%	0.052	0.002	10.7%	0.070	0.002	10.9%	0.089	0.003	14.5%	0.004	0.000

Notes: α_{ve} denotes significant levels of VaR and ES. Three columns of each model forecast results reports coverage ratio (left), p-value of DQ test (middle) and p-value of DES test respectively.

Table 3.20: Priors Specification

Parameters	Prior Distribution
μ	$N(0,1)$
κ	$N(0,1)\mathbf{1}_{\kappa>0}$
θ	$N(0,1)\mathbf{1}_{\theta>0}$
σ_V	$IG(2.5,0.1)$
ρ	$U(-1,1)$
ξ^{Y+}	$N(0,50)\mathbf{1}_{\xi^{Y+}>0}$
ξ^{Y-}	$N(0,50)\mathbf{1}_{\xi^{Y-}<0}$
ξ^{V}	$N(0,10)\mathbf{1}_{\xi^{V}>0}$
$\boldsymbol{\vartheta}$	$N(0,0.1)\mathbf{1}_{\boldsymbol{\vartheta}>0}$
λ_0^{Y+}	$N(0,0.001)\mathbf{1}_{\lambda_0^{Y+}>0}$
λ_0^{Y-}	$N(0,0.001)\mathbf{1}_{\lambda_0^{Y-}>0}$
λ_0^{v}	$N(0,0.001)\mathbf{1}_{\lambda_0^{v}>0}$
β	$N(0,0.3)\mathbf{1}_{\beta>0}$
$\tilde{\alpha}$	$N(0,0.2)\mathbf{1}_{\tilde{\alpha}>0}$
$\tilde{\beta}$	$N(0,10)\mathbf{1}_{\tilde{\beta}>0}$

Notes: This table presents priors settings in MCMC algorithm of parameter estimations

Chapter 4

Forecasting Bitcoin

4.1 Introduction

As a new financial asset, Bitcoin has attracted many studies in recent years. There have been a number of studies showing that the Bitcoin market is not fully efficient (Urquhart 2016, Nadarajah & Chu 2017, Wei 2018), is strongly affected by media and investors' sentiment (Urquhart 2018, Sapkota 2022), is often associated with high volatility (Klein et al. 2018, Shen et al. 2020), has been subject to market manipulations (Gandal et al. 2018) and has been used for illegal activities (Foley et al. 2019). The literature also shows the hedging benefits of the Bitcoin (Corbet et al. 2019, Guesmi et al. 2019, Anyfantaki et al. 2021), how blockchains manage transactions (Jiang et al. 2022, Zhang et al. 2022), and has used machine learning techniques to forecast prices and place limit orders (Atsalakis et al. 2019, Schnaubelt 2022).

In addition to these features of the Bitcoin market, another notable characteristic of Bitcoin is the large number of jumps (significant discontinuities) in Bitcoin prices. Gronwald (2019) discusses the similarity of Bitcoin to commodities and shows the presence of price jumps is more extreme in the Bitcoin market than in other markets. Shen et al. (2020) show the importance of jumps in forecasting realised volatility, while Chaim & Laurini (2018) use a stochastic volatility model and show the linkage

between prices jumps and formative events in the Bitcoin market, such as hacks. In this chapter, we consider whether the jumps in Bitcoin can be used to predict Bitcoin returns.

Earlier studies that allow jumps in continuous-time asset pricing models often assumed a relatively simple structure for the jump components, namely, that they are serially independent. They often model jumps by compound Poisson processes or Lévy processes (see e.g., Merton 1976, Bates 1996, Duffie et al. 2000, Eraker 2004). In the recent decade, a new dependence structure become popular. The dependence structure proposed by Hawkes (1971*a*,*b*) allows jump components in one dimension[1] to raise the probability of future jumps in both its own dimension and other dimensions. This is often referred to as 'self-excitation' and 'cross-excitation', respectively. Aït-Sahalia et al. (2015) incorporate this structure into a continuous-time semi-martingale model and relax the assumption of independent jump increments. Aït-Sahalia et al. (2015) consider the propagation of jumps in stock markets worldwide, while Fulop et al. (2015) also find evidence of self-exciting jumps, especially during financial crises. Other papers with finance applications using this dependence structure include Liu et al. (2021), Ketelbuters & Hainaut (2022).

There is a large literature documenting jump predictions. Lee (2012) links jumps in stock markets to macroeconomic variables and firm-specific news releases. It proposes a jump predictor based on these variables plus a dummy variable for the existence of past jumps. Lee & Wang (2019) define 3 different jump betas of individual foreign exchange rates against the market. They find that currencies with higher negative jump betas can earn higher carry trade returns. Additionally, Lee & Wang (2020) propose a jump intensity regression model that includes information releases, time-of-day effects and a dummy variable for jump clustering effects. They assess the ability of their model to predict future jumps by looking at its carry trade

[1]In this chapter, we specify two dimensions of jumps the model - positive return jumps and negative return jumps. See details in the model specification section.

performance. Further, Novotnỳ et al. (2015) present evidence of high-frequency price jumps clustering in foreign exchange markets. They detect price jumps non-parametrically and propose a trading strategy to enter the position immediately after spotting a jump. They show the profitability of the strategy.

In short, there is evidence that jump increments are not necessarily serially independent, and jumps in asset prices may exhibit clustering. Given the large number of jumps in the Bitcoin market, we ask whether jumps cluster, and are predictable. If so, to what extent can we predict jumps? Are we able tp predict the direction of jumps, in addition to their arrival? Finally, to what extent does predicting jumps help predict Bitcoin returns? These are the main questions we aim to answer in this chapter.

In this chapter, we embed a mutually-exciting point process in a stochastic volatility model with jumps, where we separate out the positive and negative return jumps and model their underlying intensity differently. We estimate the model using a Bayesian Markov chain Monte Carlo framework. In the out-of-sample analysis, we fix static parameters at posterior means and use a particle filter to estimate latent variables at each state. Then, we calculate one-step-ahead forecasts of the underlying intensities (probabilities) of jumps. We assess our forecasts by the continuous ranked probability score (CRPS), where we consider two different approaches to determine whether jumps have occurred. To the best of our knowledge, we are the first to conduct probabilistic forecasting on jumps in asset returns in this way, which is our novel methodological contribution.

In addition, we develop a statistic which is the difference between the predicted intensities of positive and negative jumps. Values in the tails of this statistic indicate the probability of a positive or negative jumps is 'high', where high is relative to the distribution of the statistic. We consider whether the tail values are able to predict returns, as a complementary way of evaluating the forecasts of the probabilities of jumps (relative to CRPS), and of the value of predicting jumps for return prediction more generally. We find some explanatory power, which varies at different data

frequencies.

In our empirical analysis, we apply 5-minute Bitcoin data and a range of other frequencies up to 120 minutes and in the in-sample estimation, we find self/cross excitation behaviours diminish with the increase of data frequency. We also observe that positive and negative jumps impact one another in an asymmetric manner. In the out-of-sample analysis, we find that distributions of intensity differences[2] between positive and negative jumps have evident tails. Positive and negative tails indicate periods when probabilities of positive and negative jumps are higher and we show that on the right (left) tails, there exists more large positive (positive) returns than negative (positive) returns. Additionally we find returns are statistically significantly positive (negative) when the probabilities of positive (negative) jumps are higher and that this significance reduces with the decrease of data frequencies.

Another novel contribution of our study is the proposal of a trading strategy - to long the positive tails and short the negative tails of the intensity difference. Our trading strategy yields a Sharpe ratio of 4.36 inclusive of transaction costs, while the buy and hold (BaH) portfolio is around 1.4. However, our trading strategy shows little difference from the BaH portfolio when the data frequency is lower than 45 minutes. We also show that the average simple return of our strategy is not significantly higher than that of the BaH portfolio; however, our maximum drawdown (MDD) is much lower. This can indicate the ability of our strategy to mitigate tail risks. Our jump prediction approach also provides potential for risk management, especially during a financial crisis when the assumption of independent jumps is no longer robust. We also conduct a further robustness check on our trading strategy. We consider a range of trading costs, trading gaps and performance evaluation ratios. We also examine whether our strategy performs better during: high/low volatility periods; during periods of price increases/decreases; and during periods of high/low uncertainty for the cryptocurrency market.

[2]Note that intensities here are predicted intensities in the out-of-sample.

Past literature forecasts jumps in asset returns primarily based on information releases (see e.g. Lee & Wang 2019, 2020)[3]. In this chapter, we focus more on probabilistic forecasting of return jumps based on the jump clustering effects. This marks out our approach from the literature. We set up a parametric model to quantify the probability of jump occurrences, and conduct probabilistic forecasting on future return jumps. The advantage of our approach is the ability to quantify and predict the probability of future jumps without relying on extraneous variables such as information releases.

The rest of this chapter is organised as follows. Section 2 introduces our model. Section 3 presents in-sample estimation and out-of-sample prediction approaches. Section 4 presents our empirical results. Section 5 introduces a trading strategy. Section 6 concludes the chapter. Some technical results are confined to the appendix.

4.2 Model Specifications

In this section, we introduce specifications of return and variance process and jump components with mutually-exciting intensities.

4.2.1 Return and Variance Process

We model Bitcoin returns (y_t) as differences of natural logarithm of Bitcoin prices (p_t), $y_t = log(p_t) - log(p_{t-1})$. We let Bitcoin returns and corresponding variance (v_t) evolve via:

$$y_t = \mu\Delta + \sqrt{v_{t-1}\Delta}\epsilon_t^y + \xi_t^+ J_t^+ + \xi_t^- J_t^-, \tag{4.1}$$

$$v_t = v_{t-1} + \kappa(\theta - v_{t-1})\Delta + \sigma_v\sqrt{v_{t-1}\Delta}\epsilon_t^v, \tag{4.2}$$

[3]For example, Lee & Wang (2020) estimate a logit model for the jump intensity, the probability of a jump, $\lambda_t = 1/\left(1 + exp(-\theta_0 - \sum_{j=1}^{7}\theta_j X_{j,t} - \sum_{h=0}^{22}\delta_h T_{h,t} - \gamma CL_t)\right)$, where $X_{j,t}$ denotes a set of dummy variables that equal 1 when there are information releases. $\delta_h T_{h,t}$ is a time indicator of different trading hours. CL_t is another dummy variable and equals 1 when there are jumps in the past 30 minutes.

where Δ is the time unit, and we take $\Delta = \{\frac{1}{288}, \frac{1}{96}, \frac{1}{48}, \frac{1}{32}, \frac{1}{24}, \frac{1}{12}\}$ corresponding to the data frequency $\delta = \{5, 15, 30, 45, 60, 120\}$ minutes. μ is a drift term. κ and θ are the mean reversion rate, and the long-run variance mean. σ_v denotes the volatility of volatility. ϵ_t^y and ϵ_t^y are two correlated normal random variable ($\mathcal{N}(0,1)$) with $corr(dW_t^p, dW_t^v) = \rho$. $\xi_t^+ J_t^+$ and $\xi_t^- J_t^-$ presents the upward and downward return jump components with underlying intensities $P(J_t^+ = 1) = \lambda_t^+$ and $P(J_t^- = 1) = \lambda_t^-$ respectively. We let jump sizes jointly follow a normal distribution with a mean μ_j and standard deviation σ_j $\left(\xi_t^+, \xi_t^- \sim \mathcal{N}(\mu_j, \sigma_j)\right)$. The specifications, at this level, resemble the stochastic volatility with jump (SVJ) model proposed by Bates (1996). [4]

4.2.2 Mutually-Exciting Jump Process

We specify a 2-dimensional mutually-exciting jump process on the underlying intensity λ_t^+, λ_t^-:

$$\lambda_t^+ = \lambda_0^+ + \sum_{0<\tilde{t}^+<t-1} \alpha_+^+ e^{-\beta_+^+(t-\tilde{t}^+)} + \sum_{0<\tilde{t}^-<t-1} \alpha_-^+ e^{-\beta_-^+(t-\tilde{t}^-)} \tag{4.3}$$

$$\lambda_t^- = \lambda_0^- + \sum_{0<\tilde{t}^+<t-1} \alpha_+^- e^{-\beta_+^-(t-\tilde{t}^+)} + \sum_{0<\tilde{t}^-<t-1} \alpha_-^- e^{-\beta_-^-(t-\tilde{t}^-)} \tag{4.4}$$

where $\{\lambda_0^+, \lambda_0^-\}$ are baseline intensities, which are constant. $\tilde{t}^+$ and $\tilde{t}^-$ denotes time indices when there are positive and negative jumps respectively ($J_t^+ = 1$ and $J_t^- = 1$). α denotes the additional intensity produced by past jumps, β governs the decaying speed of the additional intensity α. For example, the intensity of positive jumps remains λ_0^+ during peaceful times. When positive jumps occur, they will immediately raise the intensity of positive jumps by α_+^+, but this intensity will

[4] We intentionally ignore the variance jump component in the model. We acknowledge the importance of variance jumps, however, there are evidence showing variance jumps have little impact on intensities of future return jumps (see e.g. Maneesoonthorn et al. 2017, Chen et al. 2021), on which we focus more in the chapter. Therefore, we remain the model parsimonious, which can also reduce the computational burden.

decay over time by a speed governed by β_+^+. Similarly, occurrences of negative jumps will raise the probability of positive jumps by α_-^+. Therefore, the intensity of jumps equals a baseline intensity plus the impact of all past jumps, including self-excitation and cross-excitation.

4.3 In-Sample Estimation and Out-of-Sample Filtering

In this section, we introduce our in-sample parameter estimation techniques and out-of-sample particle filtering method.

4.3.1 Bayesian Inference on Parameters

Like most literature studying stochastic volatility models, we estimate the model by a Bayesian Markov chain Monte Carlo (MCMC) framework. We denote the static parameter vector as $\Theta = \{\mu, \kappa, \theta, \sigma_v, \rho, \mu_j, \sigma_j, \boldsymbol{\lambda_0}, \boldsymbol{\alpha}, \boldsymbol{\beta}\}$, where $\boldsymbol{\lambda_0} = \{\lambda_0^+, \lambda_0^-\}$ $\boldsymbol{\alpha} = \{\alpha_+^+, \alpha_-^+, \alpha_+^-, \alpha_-^-\}$ and $\boldsymbol{\beta} = \{\beta_+^+, \beta_-^+, \beta_+^-, \beta_-^-\}$. We further the latent variable vector as $\mathcal{Z}_t = \{v_t, \xi_t^+, \xi_t^-, J_t^+, J_t^-\}$. We conduct a Bayesian inference on static parameters and latent variables. The joint posterior distribution is as follows:

$$P(\Theta, \mathcal{Z}_t | y_t) \propto P(y_t | \Theta, \mathcal{Z}_t) P(\mathcal{Z}_t | \Theta) P(\Theta) \tag{4.5}$$

We adopt a Markov Chain Monte Carlo (MCMC) method to sequentially sample from the posterior distribution. Details of the algorithm and specifications of priors can be found in Appendix C.1. We follow Lazar & Qi (2022) to construct our posterior distributions, which are provided in their online appendix.

4.3.2 Particle Filtering and Out-of-Sample Predictions

In the out-of-sample, we firstly fix the static parameters Θ at its posterior mean, $\widehat{\Theta}$. We define latent variables of interests in the out-of-sample period as $\mathcal{Z}_t^* =$

$\{v_t, J_t^+, J_t^-\}$. Then, for every time t in the out-of-sample period, we adopt a particle filter proposed by Pitt & Shephard (1999) to sample from $p\left(\mathcal{Z}_t^*|y_t, \widehat{\Theta}\right)$ and filter $\widehat{\mathcal{Z}_t^*}$. Then, we conduct a one-step ahead forecast on J_{t+1}. Specifically, for every time t, we start with N random particles of variance $v_t^{(i)}, i = 1, ..., N$, and evaluate these particles with an importance weight $w_t^{(i)}$. Then, we generate $p\left(J_t|y_t, v_t^{(i)}, \widehat{\Theta}\right)$, from which we initiate another N random samples to approximate the predictive distribution $p\left(\lambda_{t+1}|y_t, \mathcal{Z}_t^{*(i)}, \widehat{\Theta}\right)$[5]. Full details of the particle filter sampling and re-sampling scheme are provided in Appendix C.2.

We evaluate our predictive distribution by a continuous ranked probability score (CRPS), which takes the following form:

$$CRPS = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N}(\widehat{\lambda}_{t+1}^{(i),(+/-)} - J_{t+1}^{(+/-)})^2 \tag{4.6}$$

this is also equivalent to the integral of the Brier scores for the binary probability forecasts (Brier et al. 1950, Matheson & Winkler 1976, Hersbach 2000). Also see discussions on CRPS by Gneiting et al. (2007) and Gneiting & Raftery (2007). The $J_{t+1}^{(+/-)}$ in the equation denotes the actual values of positive and negative jumps, on which we forecast separately. We take two measures of them in the empirical works. We firstly estimate the whole sample by Bayesian MCMC and retrieve J_t. In addition, we also use a non-parametric jump filtering method (NP filtering) as an alternative jump estimation. The procedure of the filtering method can be seen in Appendix C.3. Also see a comprehensive review on high-frequency non-parametric jump tests by Maneesoonthorn et al. (2020).

[5]The J_t here denotes $\{J_t^+, J_t^-\}$, and λ_{t+1} denotes $\{\lambda_{t+1}^+, \lambda_{t+1}^-\}$.

4.3.3 Intensity Differences

From the predictive distribution $p\left(\lambda_{t+1}|y_t, \mathcal{Z}_t^{*(i)}, \widehat{\Theta}\right)$, point forecasts of the underlying intensities $\widehat{\lambda}_{t+1}^{+}$ and $\widehat{\lambda}_{t+1}^{-}$ are calculated by taking means of those N samples:

$$\widehat{\lambda}_{t+1}^{+} = \sum_{i=1}^{N} \lambda_{t+1}^{(i)+}, \qquad \widehat{\lambda}_{t+1}^{-} = \sum_{i=1}^{N} \lambda_{t+1}^{(i)-}, \tag{4.7}$$

which are also the one-step ahead predicted probabilities of jumps at $t+1$. Note $\widehat{\lambda}_{t+1}^{+}$ and $\widehat{\lambda}_{t+1}^{-}$ equal $\widehat{\lambda}_{0}^{+}$ and $\widehat{\lambda}_{0}^{-}$ during peaceful times, and rise to higher levels given occurrences of different jumps. However, a higher $\widehat{\lambda}_{t+1}^{+}/\widehat{\lambda}_{t+1}^{-}$ does not necessarily indicate a higher probability of positive/negative return jumps since $\widehat{\lambda}_{t+1}^{+}$ and $\widehat{\lambda}_{t+1}^{-}$ can both rise simultaneously to a very higher level. However, there can be only one direction of jump at each time point. For example, when the predicted intensities both equal $\phi\%$ ($\widehat{\lambda}_{t+1}^{+} = \widehat{\lambda}_{t+1}^{-} = \phi\%$), these two forecasts become less informative for market practitioners in predicting the direction of jumps. Surely, one could argue that we can at least know the overall probability of jumps at $t+1$ is rising; however, it is less informative without knowing the direction of jumps at $t+1$.

To address this issue, we develop a new statistic, which takes the difference between these two predicted intensities:

$$\lambda_{t+1}^{d} = \widehat{\lambda}_{t+1}^{+} - \widehat{\lambda}_{t+1}^{-} \tag{4.8}$$

The essential idea of considering this statistic is to focus on those when predicted probabilities of either positive or negative jumps are higher - instead of both of them being higher. Therefore, λ_{t+1}^{d} equals ($\widehat{\lambda}_{0}^{+}$ - $\widehat{\lambda}_{0}^{-}$) most of the time. It becomes more extreme when the probability of either positive or negative jumps is relatively higher. They are captured by two tails of distributions of λ_{t+1}^{d}.

Then, it is also of our interest to study the Bitcoin returns on the tails of λ_{t+1}^{d} distributions - whether returns lying on these tails have more extreme variations and whether jump predictions translate to return predictions in Bitcoin. We further

define:

$$S_t = \begin{cases} 1, & \text{if } \lambda_t^d > \lambda_0^+ - \lambda_0^- + C_{\mathcal{N}}^{-1}(1-k\%) \cdot \sigma(\lambda_{1:t-1}^d) > 0 \\ -1, & \text{if } \lambda_t^d < \lambda_0^+ - \lambda_0^- + C_{\mathcal{N}}^{-1}(k\%) \cdot \sigma(\lambda_{1:t-1}^d) < 0 \\ 0, & \text{elsewise} \end{cases} \tag{4.9}$$

where $C_{\mathcal{N}}^{-1}(\cdot)$ denotes an inverse cumulative normal distribution. $\sigma(\lambda_{1:t-1}^d)$ represents the standard deviation of the statistic λ^d up to time $t-1$. In the empirical studies, we set $k = \{2.5, 5, 10, 15\}$.

For the convenience of further studies, we further define $S_t^+ := \mathbb{1}_{\{S_t=1\}}$, $S_t^- := \mathbb{1}_{\{S_t=-1\}}$. Therefore, $S_t = S_t^+ - S_t^-$. The basic idea is taking tails of λ_t^d, which stands for when the probability of positive jumps is significantly higher (S_t^+), and when that of negative jumps is higher (S_t^-). It is of our interest to investigate the return distributions when $S_t^+ = 1$ and $S_t^- = 1$. Note again that these two indicators $\{S_t^+, S_t^-\}$ at time t are estimated using the information up to $t-1$. Therefore, we are examining how are returns distributed when the probability of positive/negative jumps are expected to be higher.

This evaluation can also be regarded as an alternative assessment of probabilistic forecasts on return jumps in addition to the CRPS. Specifically, if our forecasting approach is valid, there should be more upward return variations when $S_t^+ = 1$, and downward return variations when $S_t^- = 1$. Thus, returns when $S_t^+ = 1$ will be significantly higher and returns when $S_t^- = 1$ will be significantly lower. On the contrary, if the forecasting does not work, $\{S_t^+, S_t^-\}$ will not provide any indications on the probability of jumps and return distributions. Thus, returns when $S_t^+ = 1$ and $S_t^- = 1$ will not make significant differences to others.

Following this logic, we further analyse two tails of λ_t^d by a regression. We regress Bitcoin simple returns with S_t^+ and S_t^-:

$$y_t = a + b_1 S_t^+ + b_2 S_t^- + u_t \tag{4.10}$$

We also remark that this regression is valid since S_t^+ and S_t^- are predicted using information up to time $t-1$. We are interested in the significance of coefficients

$\{b_1, b_2\}$. In addition, due to the potential auto-correlation of the error term u_t in the model, we adopt a Newey–West heteroscedasticity and autocorrelation consistent estimator.

Later in the empirical section, we will propose a trading strategy, which long the Bitcoin when $S_t^+ = 1$ and short the Bitcoin when $S_t^- = 1$. Therefore, S_t act as trading signals. Regressing Bitcoin returns with trading signals can also assess the validity of the model and the strategy[6]. We note that the significance of the regression is a necessary but not sufficient condition for the model (or the strategy) being 'valuable', since returns of a strategy may be subject to certain constraints, for example, trading costs.

4.4 Empirical Application

This section introduces our empirical studies, including in-sample estimation results and out-of-sample predictions.

4.4.1 Data

We collect the Bitcoin data (BTC/USD) traded on the Bitstamp exchange from an open platform `bitcoincharts.com`. We select this exchange because it is one of the first and most liquid Bitcoin exchanges (Brandvold et al. 2015). We collect data from 1st January 2014 to 31st December 2021. Then, we re-sample the tick data by taking the last price of every 5 minutes, which is the frequency we would consider in the empirical work. We also consider a range of other frequencies for comparisons and validations. Thus, we take $\delta = \{5, 15, 30, 45, 60, 120\}$ minutes. We separate the data to an in-sample and out-of-sample period on 31st December 2018. This will make 5 years in the in-sample and 3 years in the out-of-sample period. Therefore,

[6]Moskowitz et al. (2012) conduct a similar regression analysis to assess a momentum trading strategy

the first 5-year data is used to estimate the static parameter vector $\widehat{\Theta}$, which is fixed in the 3-year out-of-sample period.

4.4.2 In-Sample Estimations

We estimate the model and plot return, return extracts jumps and volatility in Fig. 4.1. Note, return and volatility in Fig. 4.1 are all re-scaled to a daily level (e.g. 5-minutes returns are multiplied by 288, 60-minutes returns are multiplied by 24).

[INSERT FIGURE 4.1 ABOUT HERE]

Table 4.1 presents the posterior mean and standard deviation of estimated parameters in the return and variance processes. We find that the same data sampled at different frequencies reports different parameters, but they do have some patterns. For example, the drift of returns μ is close to 0 with 5-minutes data and increases with lower frequency. While the mean-reversion speed of variances κ is the highest with 5-minutes data. Additionally, the results of σ_j indicate a higher variation of jump sizes with data in higher frequencies.

[INSERT TABLE 4.1 ABOUT HERE]

Posteriors of parameters in the mutually-exciting jump intensity process are presented in Table 4.2. The results are re-scaled to a daily level. The baseline intensities, $\{\lambda_0^+, \lambda_0^-\}$, are approximately 2.5% suggesting the baseline probability of return jumps is around 2.5%. This probability does not significantly vary from data frequencies. Regarding the probability of positive return jumps raised by others ($\{\alpha_+^+, \alpha_-^+\}$), the intensity raised by past negative jumps α_-^+ is slightly higher than that raised by positive jumps α_+^+ under 5-minutes data. This pattern is especially evident under 60,120-minutes data. The same pattern can be found in the probability of negative jumps, which shows the strong impact of negative jumps. However, results of $\{\beta_+^+, \beta_-^+, \beta_-^+, \beta_-^-\}$ show a higher decay speed of probability raised by negative jumps. More importantly, we find $\{\alpha_+^+, \alpha_-^+, \alpha_-^+, \alpha_-^-\}$ remain the highest under

5-minutes data, indicating more evident jump clustering features when data frequency is higher.

[INSERT TABLE 4.2 ABOUT HERE]

Overall, in-sample estimation results show that positive and negative jumps impact one another asymmetrically. The results also imply that jump clustering is more likely to be observed in higher frequency data and diminish in longer horizons. This naturally provides us an opportunity to examine the forecasting performance of data with different extent of jump clustering.

4.4.3 Out-of-Sample Filtering Results

As is introduced in Section 4.3.2, we fix static parameters at their posterior means $\widehat{\Theta}$, and filter the latent states at each time point. Then we forecast on λ_t and then take its tails S_t. Figure 4.2 presents the histogram of λ_t^d across different frequencies. There are evident tails of the empirical λ_t^d distributions, especially when the frequency exceeds 30 minutes. Table 4.3 reports the number of observations, J_t and S_t, under different data frequency δ(min) and tails parameters $k\%$. The number of jumps filtered out in the out-of-sample takes 3-5% of total observations.

[INSERT FIGURE 4.2 ABOUT HERE]

[INSERT TABLE 4.3 ABOUT HERE]

We forecast the underlying intensities $\widehat{\lambda}_{t+1}^{+}$ and $\widehat{\lambda}_{t+1}^{-}$, and corresponding CRPS is reported in Table 4.4. The top panel use J_t estimated by the Bayesian method as the actual value, and the bottom panel use J_t estimated by the NP filtering as the actual value. The $\mathcal{M}_1$ denotes $\widehat{\lambda}_{t+1}$ predicted by our model. We benchmark two forecasts. The $\mathcal{M}_2$ takes $\widehat{\lambda}_{t+1} = \sum(J_t)/T$. where T denotes number of observations. This setting assumes J_t arrives under a Poisson process, whose arrivals are serially independent with a constant intensity. $\mathcal{M}_3$ takes $\widehat{\lambda}_{t+1} = 0$, which simply assumes

there are no jumps. We also report results of a Diebold-Mariano (DM) test (Diebold & Mariano 2002) in Table 4.4. Values in brackets are p-value of the DM test that tests forecasts by $\mathcal{M}_1$ against other models. Note that in this test, we calculate the probability score as follow:

$$Score_t = \frac{1}{N}\sum_{i=1}^{N}(\widehat{\lambda}_{t+1}^{(i),(+/-)} - J_{t+1}^{(+/-)})^2 \tag{4.11}$$

such that it becomes a time-series score to be applied in the test[7].

[INSERT TABLE 4.4 ABOUT HERE]

CRPS measures the extent to which our forecasts calibrate to the arrivals of jumps. A lower value of CRPS indicates a better calibration of the forecasts. As is shown in Table 4.4, with 5-minutes data, forecasts from our model ($\mathcal{M}_1$) reports the lowest CRPS. Results of the DM test suggest that our model consistently outperform others when the data frequency is higher than 30 minutes, while they are indifferent to forecasts from other benchmark models under lower frequency data.

These results are in line with in-sample estimations - self-exciting jumps are observed more evident under higher frequency data with higher value of $\{\alpha_+^+, \alpha_-^+, \alpha_-^+, \alpha_-^-\}$, which translates to better performance in return jump predictions. The CRPS results also show the importance of considering jump self-excitation in the model since modelling and predicting return jumps by the mutually-exciting process is clearly superior to processes assuming independent or no jumps.

4.4.4 Returns on Tails of λ_t^d Distribution

We identify left and right tails of λ_t^d distribution (S_t^+ and S_t^-) according to methods we discussed in Section 4.3.3. We also highlight that S_t^+ and S_t^- can also be

[7]For a specific example, we have two time-series score (s_t^1 and s_t^2) from $\mathcal{M}_1$ and $\mathcal{M}_2$. We calculate Diebold-Mariano statistic as $\text{DM} = \bar{s}/\left(\sqrt{(\gamma_0 + 2\sum_{k=1}^{h-1}\gamma_k)/n}\right)$, where $\bar{s} = \mathbb{E}(s_t^1 - s_t^2)$ and γ_k denotes the autocovariance of $s_t^1 - s_t^2$ at lag k. n denotes number of forecasts and we take $h = n^{\frac{1}{3}} + 1$.

treated as trading signals. We firstly evaluate returns identified by these trading signals (lying on tails of λ_t^d distributions). The essential idea is to look at how is return distributed when the probability of positive/negative jumps is significantly higher. In Table 4.5 and 4.6, we summarise the percentage of return sizes that are greater than $\{0, \sigma(y_t), 1.5\sigma(y_t), 2\sigma(y_t)\}$, and return sizes that are smaller than $\{0, -\sigma(y_t), -1.5\sigma(y_t), -2\sigma(y_t)\}$. The $\sigma(y_t)$ denotes the standard deviation of returns of the whole sample[8].

The results show that the tails of λ_t^d have strong indications of extreme returns. For example, on the right tails of λ_t^d under 5-minutes data ($\delta = 5, k = 10$) (Table 4.5, second row), 21.8% of returns of the right tail is greater than 1 standard deviation of returns, and 8.9% of that are greater than 2 standard deviations of returns. By contrast, only 6.7% of returns on the right tail are smaller than negative 1 standard deviation of returns, and 2.7% of that is smaller than negative 2 standard deviations of returns. Vice versa, we can find similar patterns on the left tail, which is reported in Table 4.6. However, looking at results under other frequency data, this pattern diminishes with the decrease of data frequencies. Under 120-minutes data, the percentage of returns lying on the right tail is not significantly distinct from that on the left tail.

[INSERT TABLE 4.5 ABOUT HERE]

[INSERT TABLE 4.6 ABOUT HERE]

These results strongly indicate the predictability of λ_t^d's tails on future returns. If our approach is invalid, the return distributions on those tails should resemble those under 120-minutes data, where the distribution of extreme returns tends to be symmetric regardless of predicted probabilities of jumps.

Regression results of (4.10) are presented in Table 4.7. We re-scale the Bitcoin return data to the daily level as the regression input. The regression results confirm

[8]We tried using the standard deviation of returns in the in-sample period, and returns up to $t-1$. Results do not vary too much.

that returns on the λ_t^d's right tail (S_t^+) are significantly higher, and those on the left tail are significantly lower. This suggests the predictability of S_t on Bitcoin returns. However, with the decrease in data frequencies, the coefficient estimates reduce from ± 0.06 to ± 0.001 approximately. The corresponding test statistics also reduce from ± 12 to ± 0.5 approximately.

[INSERT TABLE 4.7 ABOUT HERE]

4.5 Trading on Jump Clustering

Based on these results, we form a portfolio that goes long on the right tail and short the left tail of λ_t^d distribution. Specifically, we long the BTC/USD when $S_t^+ = 1$ and short the BTC/USD when $S_t^+ = 1$ by a market order. The strategy is valid since S_t are generated based on information up to time $t-1$. We also consider a long-only strategy due to the short-selling constraint of the Bitcoin. This section will introduce trading costs considered in our strategy and performance evaluations.

4.5.1 Trading Costs

We assume a transaction cost of 0.15% for each transaction[9] (buy/sell). We also add an additional 0.05% as a liquidity cost. This number is roughly estimated by limit order book data. Since the full limit order book data of Bitcoin is not available, we collect real-time limit order book data by using the API provided by Bitstamp Exchange. We collect data every hour from 15/03/2022 to 15/04/2022. We find that executing a market order worth \$20,000[10] maximally reduces the return by 0.05%. This is calculated by the percentage difference between the weighted average of

[9]Bitstamp Exchange has a range of transaction fee schedules. The fee varies according to account trading volume in the past 30 days. We take 0.15% when the account trading volume is around \$600,000.

[10]We use total trading volume in dollar divided by number of transactions in 2021 and get the average trading volume (\$) per transaction equals to approximately \$20000.

market order price (assume the order is filled), and the spot market price. Since we intend to implement a high-frequency trading strategy, 30 ($600,000/$20,000) trades will be sufficient to fulfil the monthly $600,000 trading volume requirements. Thus, it is safe to assume a 0.15%+0.05% transaction cost structure in assessing our trading strategy. Brauneis et al. (2021) construct high-frequency measures of transaction costs and liquidity costs. Their studies also support our trading cost assumptions of 0.2% being reasonable.

We also give a 15 seconds gap to execute the order. Therefore, when a return comes in each δ, we estimate v_t and J_t by particle filtering, then forecast on λ^d_{t+1} and S_{t+1}. Then, we trade the BTC/USD based on S_{t+1} on the 15th second of the next minute. Running a particle filter and forecasting takes less than 2 seconds. Executing a market order takes less than 5 seconds. Therefore, a 15-seconds gap should be enough for the algorithmic trading.

4.5.2 Performance Evaluation

In Fig. 4.3, we plot cumulative simple return of the trading strategy. Since short selling in the Bitcoin market was not available throughout our sample and is costly, we also consider a long-only trading strategy. Fig. 4.4 plots the cumulative simple return of the strategy. In these two figures, the green line presents the cumulative return of the BaH strategy, while the other four lines plot that of our strategy with different k%. These figures show that the cumulative return tends to be less volatile than the BaH return. Furthermore, our strategy tends to be less effective when the data frequency is lower than 45 minutes, which is in line with the in-sample estimation results in Section 4.4.2 that α become smaller with lower frequency data. Also, the tails of λ^d_t distribution are less evident with lower frequency data (see Fig. 4.2).

[INSERT FIGURE 4.3 ABOUT HERE]

[INSERT FIGURE 4.4 ABOUT HERE]

In addition, we report annualised mean return (AMR), Sharpe ratio (SR), the standard deviation of returns and maximum drawdown (MDD) in Table 4.8, 4.9 and 4.10, respectively. Note all of these values are annualised. Sharpe ratio and the maximum drawdown are calculated as follows:

$$SR = \frac{\overline{y_t^p - r_t}}{\sigma(y_t^p)} \qquad MDD = max\left(\frac{p_t - max(p_{1:t-1})}{max(p_{1:t-1})}\right) \tag{4.12}$$

where r_t denotes the risk-free rate at time t, and $\overline{y_t^p - r_t}$ presents the strategy excess return at time t. Therefore, SR measures the average excess return bearing one standard deviation unit. p_t denotes the Bitcoin spot price, and MDD measures the maximum drop from peak to trough of an Bitcoin price.

To examine whether the Sharpe ratios of our trading strategies are significantly better than those of a simple buy-and-hold trading strategy, we use Ledoit & Wolf (2008)'s studentised circular block bootstrap. We set the block length as 5 and the bootstrap iterations as 1000. We perform a one-sided test with a null hypothesis that the Sharpe ratio of a trading strategy i is greater than that of a buy-and-hold strategy ($H_0 : \widehat{SR}_i - \widehat{SR}_{BaH} > 0$). The results are also reported in Table 4.9.

[INSERT TABLE 4.8 ABOUT HERE]

[INSERT TABLE 4.9 ABOUT HERE]

[INSERT TABLE 4.10 ABOUT HERE]

We find the AMR of the strategy is strongly affected by transaction costs under 5-minutes data. After-cost AMR of the strategy is not higher than BaH portfolio. However, our strategy reports outstanding performance in terms of the Sharpe ratio. Sharpe ratio of the long-only portfolio reaches 4.36 with transaction costs. However, SR of the strategy is less different from BaH SR when the frequency is lower than 45 minutes. We find that it is the low volatility of returns that lead to the high SR. As shown in Table 4.10, our strategy reports a significantly lower standard deviation and MDD. This indicates the superiority of our approach in mitigating tail risks caused by jump propagation.

4.5.3 Robustness of Trading Strategies

We conduct several robustness checks on our trading strategies. First, we calculate Calmar ratio (CR), Sortino ratio (SR), and Omega ratio (OR), which are specified as follow:

$$CR = \frac{AMR}{MDD}, \qquad SR = \frac{\overline{y_t^p - r_t}}{\sigma(y_t^p \mathbb{1}_{y_t^p<0})}, \qquad OR = \frac{\sum y_t^p \mathbb{1}_{y_t^p>0}}{\sum y_t^p \mathbb{1}_{y_t^p<0}} \tag{4.13}$$

CR and SR measure portfolio returns compensated by downside risks (MDD and downside standard deviations). OR measures positive cumulative returns against negative ones. The results of these ratios are reported in Table 4.11. Compared to a BaH portfolio, our trading strategy performs better with higher data frequencies, which is in line with previous results.

[INSERT TABLE 4.11 ABOUT HERE]

Second, we examine the strategy performance with a range of trading costs from 0.2% to 0.4%. For simplicity, we set $k = 10\%$ and only consider data frequencies higher than 45 minutes. We plot the cumulative simple return of a long-short strategy and that of a long-only strategy in Figure 4.5. We also calculate AMR and SR of portfolio returns across different trading costs and report the results in Table 4.12. We find higher frequency data are more sensitive to increasing trading costs, and the long-short trading strategy is more likely affected by higher trading costs. We also highlight that the 0.2% trading cost is already a very conservative estimates, as mentioned in Section 4.5.1. However, looking at Sharpe ratios, the strategy still works after adding an additional 0.1% on top of the 0.2%.

[INSERT FIGURE 4.5 ABOUT HERE]

[INSERT TABLE 4.12 ABOUT HERE]

Third, we examine the strategy with different trading gaps from 15 seconds to 60 seconds. Figure 4.6 plots the strategy cumulative returns with these trading

gaps. We find higher frequency data are more likely affected by longer trading gaps, especially under 5-minutes data. The impact of longer trading gaps is much smaller under 15-minutes data, and negligible under lower frequency data.

[INSERT FIGURE 4.6 ABOUT HERE]

Fourth, we conduct a further regression analysis to examine whether our strategy returns are significantly higher/lower when Bitcoin volatility is higher/lower and when Bitcoin price increases/decreases. Our regression models take the following forms:

$$y_t^p = c_0 + c_1 Q_1 + c_1 Q_2 + c_1 Q_3 + u_t, \tag{4.14}$$

$$y_t^p = d_0 + d_1 D_t + v_t, \tag{4.15}$$

where Q_1 is a dummy variable, which equals 1 when the volatility (v_t) lies in its first quartile. Similarly, Q_2 and Q_3 capture the second and the third quartile of the volatility, respectively. We take quartiles of volatility, which is estimated by Bayesian MCMC, as the regressors. Additionally, D_t equals 1 when Bitcoin return (y_t) is positive. Table 4.13 presents the regression estimates and their corresponding t-statistics. The results show that strategy returns are less likely related to volatility (see Panel A). Regression with return signs (Panel B) presents an interesting pattern. The coefficient d_1 are overall significantly positive, especially with lower frequency data. This indicates that our strategy works better when Bitcoin price increases. However, this pattern becomes less significant with higher frequency data.

[INSERT TABLE 4.13 ABOUT HERE]

Fifth, we regress our strategy returns with a cryptocurrency policy uncertainty index (CPIPolicy) and a cryptocurrency price uncertainty index (CPIPrice) proposed by Lucey et al. (2022). Our regression models read:

$$y_t^p = e_0 + e_1 CPIPolicy + w_t, \tag{4.16}$$

$$y_t^p = f_0 + f_1 CPIPrice + z_t, \tag{4.17}$$

These two indices[11] capture uncertainties caused by policy and regulatory debates and media's attention. We aggregate our strategy returns to weekly returns since indices are weekly-based. The regression results are reported in Table 4.14. Panel C and D report the regression with CPIPolicy and CPIPrice, respectively. These two regressions report similar results, in which most coefficient estimates (e_1 and f_1) are insignificant, suggesting our strategy returns are less likely relevant to the uncertainty of the cryptocurrency market.

[INSERT TABLE 4.14 ABOUT HERE]

4.6 Conclusion

In this chapter, we study on the predictability of Bitcoin returns and whether predicting jumps helps forecast returns. We propose an approach to predict jumps under mutually-exciting jump processes. We use a stochastic volatility model with self/cross exciting jump components, estimate the model using a Bayesian MCMC in the in-sample period, and iteratively estimate the latent states by a particle filter in the out-of-sample period. Then we forecast the one-step-ahead underlying intensities (probabilities) of return jumps.

We assess the forecast performance by CRSP. In its calculation, we consider jumps estimated by the Bayesian method and also estimated non-parametrically, as the actual values. Our model reports better performance against our two benchmarks. We also develop a statistic which takes the difference between positive jumps and negative jumps probabilities (λ_t^d). The intuition of considering this statistic is that its tail should be associated with significantly positive/negative returns if our prediction approach is valid. Therefore, we regress asset returns on the tails of the probability difference to investigate if Bitcoin jump predictions contribute to

[11] Indices are available on the website `https://sites.google.com/view/cryptocurrency-indices/home?authuser=0`

forecasting its return.

In the empirical analysis, we use Bitcoin data with a range of frequencies. We find patterns across data frequencies, for example, higher frequency data shows a stronger mutually-exciting (jump clustering) feature and better forecast performance in terms of the CRPS. Through the regression analysis, we also show that the tails of λ_t^d are strong indicators of returns. Specifically, returns on the positive tail are more likely positive, and vice versa.

We propose a trading strategy based on this structure. We consider transaction costs and liquidity costs in the strategy. Our trading strategy yields a Sharpe ratio of 4.36 with these costs. We find the exceptional performance stems from lower standard deviation and maximum drawdown than a buy and hold strategy. Additionally, our trading strategy becomes less effective when data frequency is lower than 45 minutes. On top of this, we conduct a robustness check of the strategy. We find that the strategy applied in higher frequency data is more sensitive to trading costs and trading gaps. Although 5-minutes data reports the best performance, it tends to deteriorate with increasing trading costs and trading gaps. We also find our strategy is less likely related to high/low volatility periods, but it significantly performs better during periods increasing price.

Our research provides financial practitioners, especially Bitcoin traders, with an approach to predict and trade on Bitcoin based on the its jump clustering behaviours. The standard deviation and maximum drawdown of our trading portfolio show that our approach is especially effective in mitigating downside risks. We also highlight that our forecasting framework has the potential to be applied to other assets exhibiting jump clustering.

C Appendices for Chapter 4

C.1 Bayesian MCMC Algorithm and Specification of Priors

We adopt a Bayesian MCMC algorithm to obtain the joint posterior distribution in Equation 4.5. We randomly draw samples from conditional posterior distributions. Our sampling algorithm is as follows:

Algorithm 4 Bayesian MCMC Algorithm

For $i = 1 : 100,000$:

1: Sample k static parameters:

Draw $\Theta_1^{(i)}$ from $p\left(\Theta_1^{(i)}|y_t, \Theta_2^{(i-1)}, \Theta_3^{(i-1)}, ..., \Theta_k^{(i-1)}, \xi_t^{+(i-1)}, \xi_t^{-(i-1)}\right)$,

$\vdots$

Draw $\Theta_k^{(i)}$ from $p\left(\Theta_k^{(i)}|y_t, \Theta_1^{(i-1)}, \Theta_2^{(i-1)}, ..., \Theta_{k-1}^{(i-1)}, \xi^{+(i-1)}, \xi^{-(i-1)}\right)$

2: Sample jump sizes

for $t = 1, 2, ..., T$:

Draw $\xi_t^{+(i)}$ from $p\left(\xi_t^{+(i)}|y_t, \Theta^{(i)}, \xi_t^{-(i-1)}\right)$,

Draw $\xi_t^{-(i)}$ from $p\left(\xi_t^{-(i)}|y_t, \Theta^{(i)}, \xi_t^{+(i)}\right)$,

3: Sample variance:

for t = 1,2, ... , T:

Draw V_t from $p\left(V_t^{(i)}|y_t, \Theta^{(i)}, \xi^{+(i)}, \xi^{-(i)}\right)$,

We ran our MCMC algorithm for 100,000 iterations, and the first 50,000 is regarded as a burn-in period. We sample the posterior distribution by a hybrid of Gibbs sampling and Metropolis-Hastings. Gibbs sampling methods are used when conjugate priors of conditional posteriors are available. For others whose posteriors are not available in closed form, we adopt the Metropolis-Hastings sampling method. Details of posterior distributions are provided by Stroud & Johannes (2014), Lazar & Qi (2022) and Rasmussen (2013). Our specification of priors are as follow:

[INSERT TABLE 4.15 ABOUT HERE]

C.2 Particle Filter

We follow Stroud & Johannes (2014), who adopt a particle filter to sample latent variables in the out-of-sample period. The particle filtering method is given by Pitt & Shephard (1999). Creal (2012) also gives a review on this technique. Our ultimate goal is to sample from $p\left(\mathcal{Z}_t^*|y_t,\widehat{\Theta}\right)$ to obtain estimates of latent variables at each time t, where $\mathcal{Z}_t^* = \{v_t, J_t^+, J_t^-\}$. We remark that we ignore sampling jump sizes $\{\xi_t^+, \xi_t^-\}$ since they are less relevant to forward simulating jump intensities $\{\lambda_t^+, \lambda_t^-\}$. We adopt a sequential importance sampling with a re-sampling scheme:

Algorithm 5 Sequential Importance Sampling with Re-sampling

At $t = 0$, for $i = 1, ..., N$

Draw $\mathcal{Z}_0^*(i) \sim g_0(\mathcal{Z}_0^*)$, set $\omega_0^{(i)} = \frac{p(\mathcal{Z}_0^*(i))}{g_0(\mathcal{Z}_0^*(i))}$

For $t = 1, ..., T$:

1: For $i = 1, ..., N$, draw $\mathcal{Z}_t^*(i) \sim g_t\left(\mathcal{Z}_t^*|\mathcal{Z}_{t-1}^*, y_t, \widehat{\Theta}\right)$.

2: Compute importance weights $\omega_t^{(i)} \propto \omega_{t-1}^{(i)} \frac{p\left(y_t|\mathcal{Z}_t^*(i),\widehat{\Theta}\right)p\left(\mathcal{Z}_t^*(i)|\mathcal{Z}_{t-1}^*(i),\widehat{\Theta}\right)}{g_t\left(\mathcal{Z}_t^*(i)|\mathcal{Z}_{t-1}^*(i),y_t,\widehat{\Theta}\right)}$.

3: Normalise importance weights $\widehat{\omega}_t^{(i)} = \frac{\omega_t^{(i)}}{\sum_{j=1}^N \omega_t^{(j)}}$.

4: Re-sample N particles by $\{\widehat{\omega}_t^{(i)}\}_{i=1}^N$ and reset $\omega_t^{(i)} = \frac{1}{N}$.

See more details of particle filtering in Creal (2012).

C.3 Non-Parametric Jump Filtering

Our jump filtering method is mainly based on Mancini et al. (2015) and Figueroa-López & Mancini (2019). We identify a jump at time t, $J_t = 1$, when the squared return is greater than a threshold, $y_t^2 > \widehat{v_t^2} \cdot 2\Delta log\frac{1}{\Delta}$. $\widehat{v_t^2}$ is a non-parametric estimator of spot variance based on pre-truncated returns:

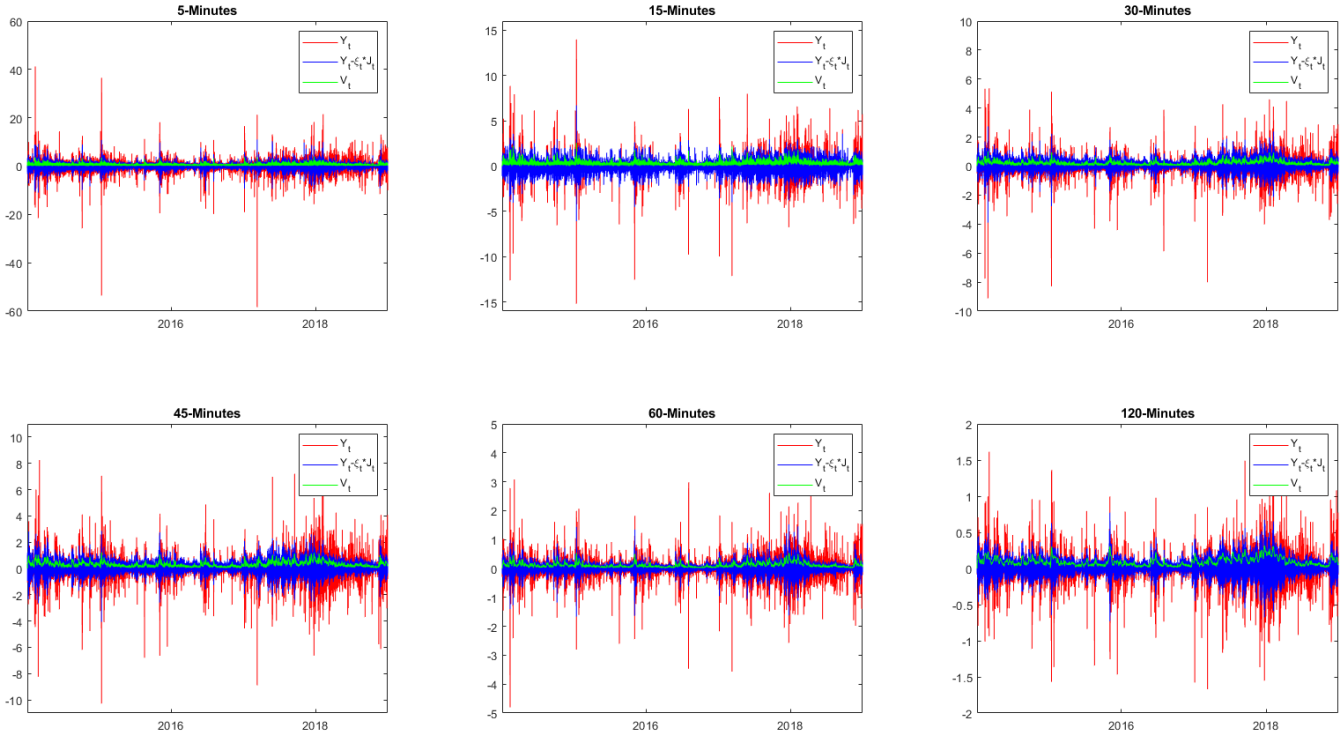
$$\widehat{v_t^2} = \sum_{\tilde{t}=1}^{n} f_h(t-\tilde{t}) y_t^2 \cdot \mathbb{1}_{\{y_t^2 \leq 9\Delta^{0.99}\}}$$

where $f_h(\cdot)$ is weight function, $f_h(t) = \frac{1}{h} \cdot \frac{e^{-|t/h|}}{2}$ with a bandwidth $h = 200\Delta$ for simplicity. The idea of this filtering is to extract those standardised squared returns

$(y_t\Delta/\widehat{v_t^2})$ which is not generated by a Brownian motion, whose absolute value is greater than the threshold $\sqrt{2log(1/\Delta)}$.

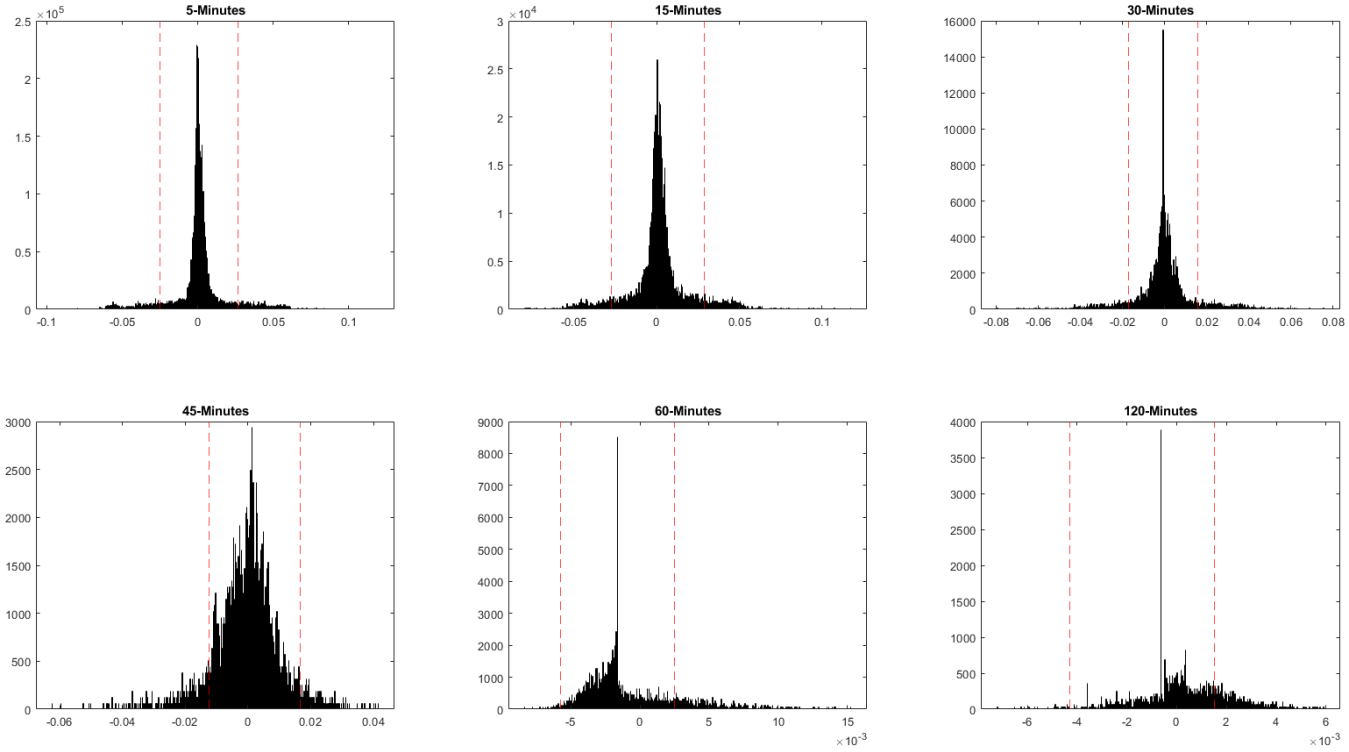
4.7 Figures and Tables for Chapter 4

Figure 4.1: Return and Volatility across Different Frequencies



Notes: In this figure, we plot return (red), return minus jumps (blue) and volatility (green) of Bitcoin across different frequency. Note, they are all re-scaled to a daily level (e.g. 5-minutes returns are multiplied by 288, 60-minutes returns are multiplied by 24). The data period is from 1/1/2014 to 31/12/2018.

Figure 4.2: Histogram of λ_t^d with $k\% = 10\%$



Notes: The data used in calculating λ_t^d is Bitcoin data from 1/1/2014 to 31/12/2018.

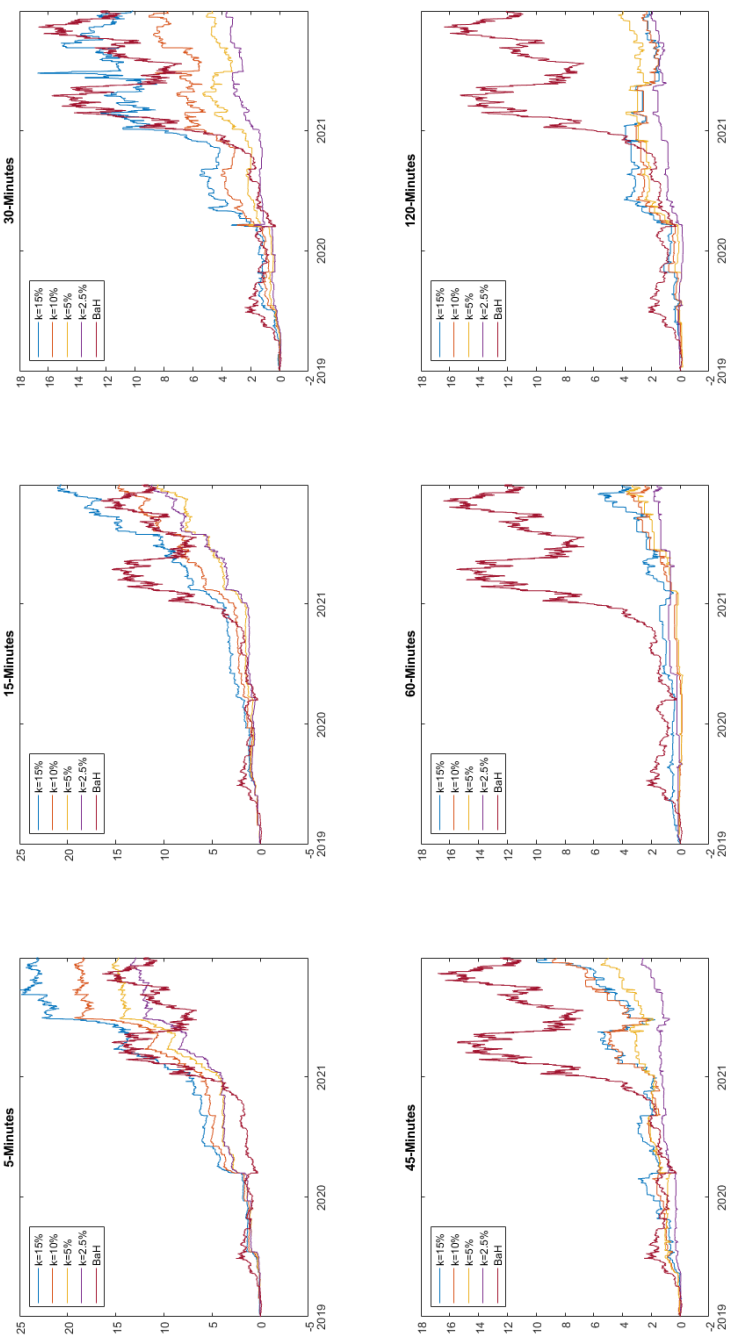


Figure 4.3: Cumulative Simple Return of Long and Short Strategy

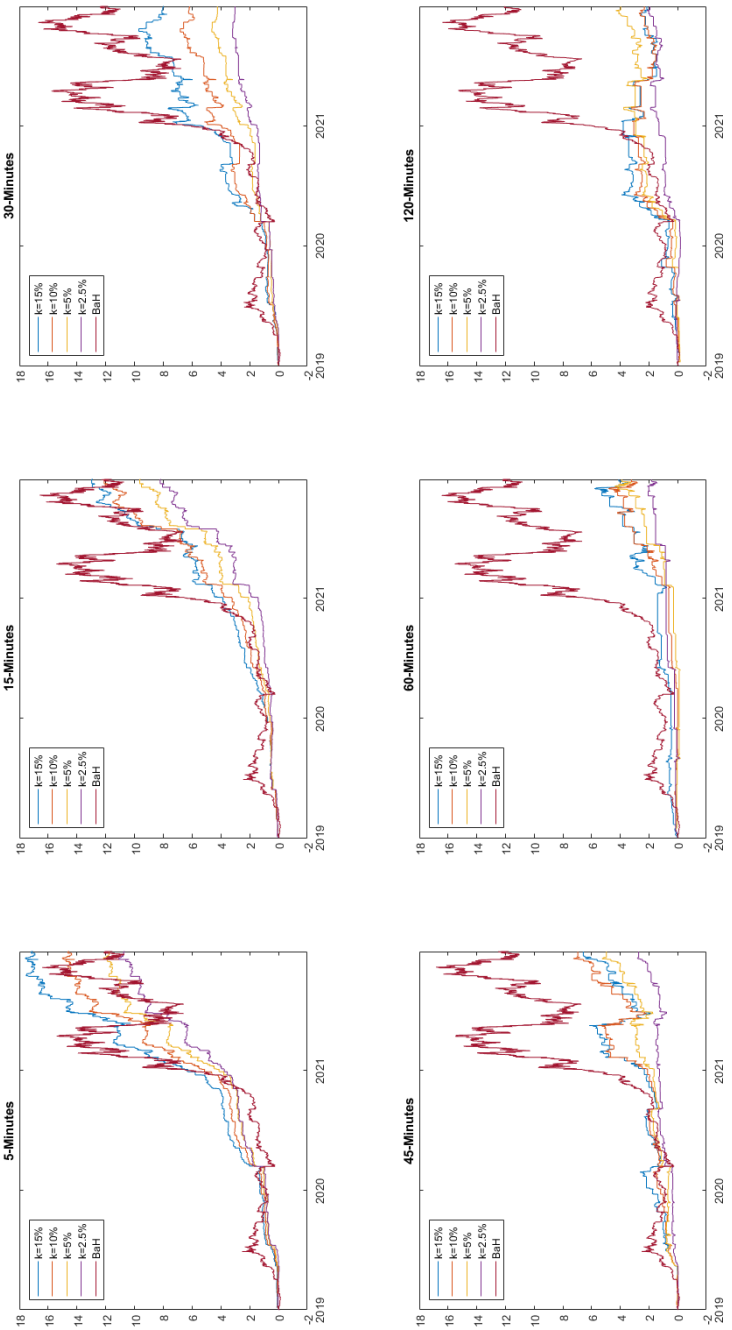


Figure 4.4: Cumulative Simple Return of Long Only Strategy

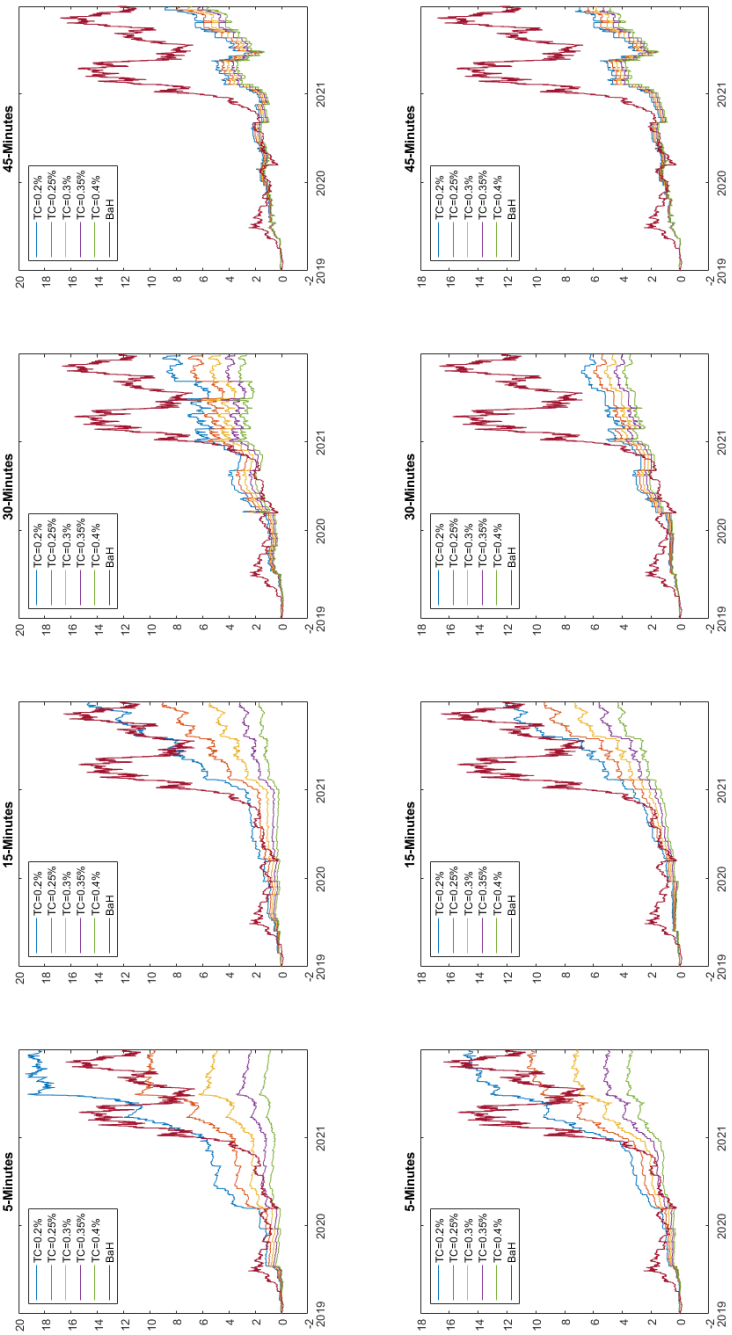


Figure 4.5: Cumulative Simple Return with Different Trading Costs

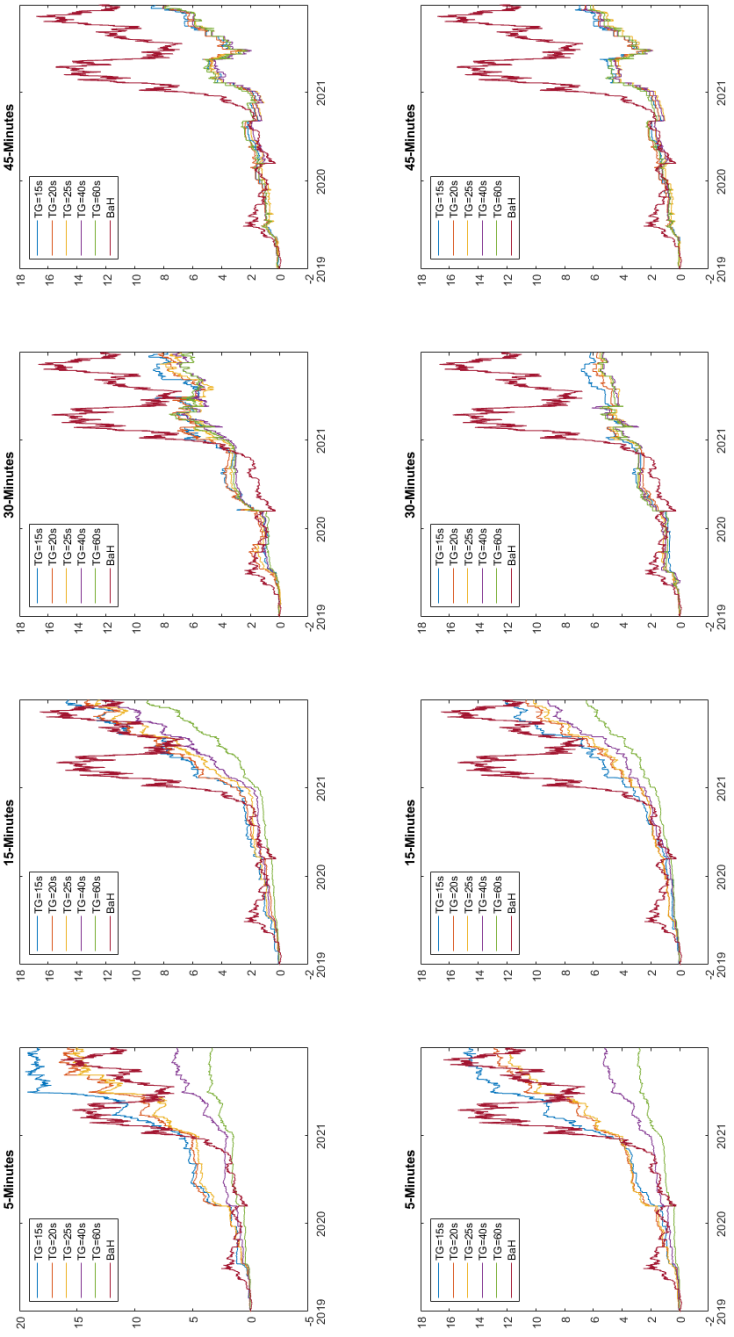


Figure 4.6: Cumulative Simple Return with Different Trading Gaps

Table 4.1: Estimated Static Parameters in Return and Variance Processes

δ (min)		μ ($*10^{-2}$)	κ	θ	σ_V	ρ	μ_J	σ_J
5	Posterior Mean	-0.02	0.116	0.119	0.182	-0.13	-0.010	2.836
	Posterior Std Dev.	0.05	0.003	0.015	0.013	0.01	0.028	0.565
15	Posterior Mean	0.06	0.092	0.160	0.067	-0.14	-0.037	1.862
	Posterior Std Dev.	0.03	0.003	0.027	0.000	0.01	0.023	0.053
30	Posterior Mean	0.07	0.014	0.165	0.032	-0.13	-0.010	1.349
	Posterior Std Dev.	0.03	0.002	0.025	0.002	0.01	0.033	0.119
45	Posterior Mean	0.11	0.011	0.145	0.043	-0.12	-0.011	1.211
	Posterior Std Dev.	0.07	0.001	0.008	0.001	0.02	0.036	0.101
60	Posterior Mean	0.16	0.019	0.192	0.057	-0.12	-0.019	0.969
	Posterior Std Dev.	0.03	0.009	0.016	0.003	0.01	0.020	0.275
120	Posterior Mean	0.24	0.011	0.169	0.032	-0.11	-0.023	0.403
	Posterior Std Dev.	0.07	0.003	0.018	0.004	0.03	0.041	0.107

Notes: these parameters are estimated with data re-scaled to daily levels. For example, 30-minutes data are multiplied by 48.

Table 4.2: Estimated Parameters in the Mutually-Exciting Jump Processes

δ (min)		λ_0^+	α_+^+	α_-^+	β_+^+	β_-^+	λ_0^-	α_+^-	α_-^-	β_+^-	β_-^-
5	Posterior Mean	0.023	0.067	0.075	0.109	0.124	0.026	0.075	0.082	0.126	0.143
	Posterior Std Dev.	0.005	0.018	0.013	0.070	0.013	0.006	0.012	0.010	0.020	0.030
15	Posterior Mean	0.024	0.054	0.067	0.084	0.136	0.019	0.059	0.067	0.142	0.111
	Posterior Std Dev.	0.005	0.017	0.016	0.051	0.012	0.005	0.010	0.008	0.021	0.023
30	Posterior Mean	0.029	0.041	0.042	0.090	0.112	0.026	0.032	0.057	0.150	0.094
	Posterior Std Dev.	0.006	0.012	0.014	0.036	0.012	0.003	0.012	0.015	0.023	0.018
45	Posterior Mean	0.020	0.037	0.034	0.096	0.145	0.017	0.035	0.036	0.055	0.173
	Posterior Std Dev.	0.005	0.014	0.013	0.018	0.013	0.005	0.016	0.014	0.011	0.038
60	Posterior Mean	0.024	0.014	0.024	0.106	0.129	0.020	0.016	0.027	0.078	0.173
	Posterior Std Dev.	0.006	0.009	0.015	0.051	0.027	0.006	0.009	0.018	0.012	0.053
120	Posterior Mean	0.025	0.018	0.026	0.126	0.490	0.023	0.011	0.019	0.082	0.387
	Posterior Std Dev.	0.006	0.013	0.017	0.043	0.157	0.007	0.008	0.014	0.026	0.154

Notes: this table presents posterior means and standard deviations of parameters of the mutually exciting kernel estimated using Bitcoin data across a range of data frequencies.

Table 4.3: Number of J_t and S_t Summary

$k\%$	δ (min)	S_t^+	S_t^-	δ (min)	S_t^+	S_t^-
15	5 (315360) (2.14%, 2.11%)	3.95%	4.00%	45 (35040) (1.55%, 1.51%)	3.72%	3.41%
10		3.38%	3.36%		2.81%	2.64%
5		2.87%	2.79%		1.95%	1.81%
2.5		2.54%	2.36%		1.33%	1.36%
15	15 (105120) (1.61%, 1.51%)	3.31%	2.94%	60 (26280) (1.63%, 1.47%)	7.64%	0.94%
10		2.75%	2.39%		5.82%	0.43%
5		2.12%	1.85%		3.68%	0.16%
2.5		1.67%	1.42%		2.39%	0.02%
15	30 (52560) (2.33%, 2.11%)	4.81%	2.87%	120 (13140) (1.92%, 1.92%)	7.83%	1.39%
10		3.27%	1.95%		7.44%	0.57%
5		2.27%	1.11%		6.35%	0.23%
2.5		1.51%	0.70%		4.63%	0.02%

Notes: the second and fifth column reports data frequency (δ), number of observations in the out-of-sample period and the percentages of jumps (positive and negative), respectively. The table also reports the percentage of tail indicators (S_t^+ and S_t^-) of intensity differences λ_t^d across different values of $k(\%)$.

Table 4.4: Continuous Ranked Probability Score (CRPS) of Jump Predictions and p-value of Associated DM test

		Bayesian Estimated Jumps as True					
	Model	5	15	30	45	60	120
J_t^+	$\mathcal{M}_1$	0.0180	0.0143	0.0216	0.0150	0.0155	0.0185
	$\mathcal{M}_2$	0.021**	0.0159**	0.0227**	0.0153	0.0157	0.0188
		(0.015)	(0.022)	(0.045)	(0.116)	(0.418)	(0.294)
	$\mathcal{M}_3$	0.0214***	0.0161**	0.0233**	0.0155*	0.016	0.0192
		(0.009)	(0.013)	(0.032)	(0.094)	(0.348)	(0.112)
J_t^-	$\mathcal{M}_1$	0.0175	0.0134	0.0197	0.0144	0.0145	0.0187
	$\mathcal{M}_2$	0.0206**	0.0148**	0.0206*	0.0148	0.0145	0.0188
		(0.011)	(0.031)	(0.067)	(0.217)	(0.695)	(0.711)
	$\mathcal{M}_3$	0.0211**	0.0151**	0.0211**	0.015*	0.0147	0.0192
		(0.003)	(0.015)	(0.048)	(0.085)	(0.436)	(0.169)
		Non-Parametrically Estiamted Jumps as True					
	Model	5	15	30	45	60	120
J_t^+	$\mathcal{M}_1$	0.0168	0.0119	0.0164	0.0123	0.0107	0.0201
	$\mathcal{M}_2$	0.0187**	0.0131*	0.0169	0.0124	0.0104	0.0202
		(0.027)	(0.075)	(0.191)	(0.425)	(0.548)	(0.496)
	$\mathcal{M}_3$	0.0191**	0.0132*	0.0172	0.0125	0.0105	0.0206
		(0.013)	(0.069)	(0.159)	(0.319)	(0.674)	(0.113)
J_t^-	$\mathcal{M}_1$	0.0138	0.0113	0.0154	0.0114	0.0143	0.0156
	$\mathcal{M}_2$	0.0159**	0.0124*	0.0162	0.0114	0.0143	0.0154
		(0.011)	(0.077)	(0.132)	(0.888)	(0.367)	(0.421)
	$\mathcal{M}_3$	0.0161**	0.0125*	0.0165*	0.0115	0.0145	0.0156
		(0.010)	(0.077)	(0.088)	(0.741)	(0.189)	(0.726)

Notes: This table presents CRPS of different models with different proxies of true value of jumps. Values in brackets reports p-value of the DM test that tests forecasts by $\mathcal{M}_1$ against other models. See the calculation of test statistics in footnote 6.
***,**,* denote the result of a DM test being significant at 1%, 5%, 10%, respectively.

Table 4.5: S_t^+ Evaluations

δ (min)	k(%)	y_t>0	y_t>$\sigma(y_t)$	y_t>1.5$\sigma(y_t)$	y_t>2$\sigma(y_t)$	y_t<-$\sigma(y_t)$	y_t<-1.5$\sigma(y_t)$	y_t<-2$\sigma(y_t)$
5	15	63.3%	20.5%	11.9%	8.5%	6.2%	3.2%	2.5%
	10	64.0%	21.8%	12.6%	8.9%	6.7%	3.4%	2.7%
	5	63.9%	22.8%	12.9%	9.2%	6.6%	3.6%	3.2%
	2.5	63.9%	22.9%	12.8%	8.7%	6.2%	3.1%	2.8%
15	15	58.9%	18.6%	8.9%	6.8%	6.6%	2.6%	1.5%
	10	59.9%	19.8%	9.7%	7.5%	6.6%	2.6%	1.4%
	5	61.2%	20.5%	10.3%	8.3%	5.8%	2.5%	1.3%
	2.5	64.5%	22.3%	11.4%	9.3%	5.5%	2.5%	1.6%
30	15	58.2%	15.7%	7.0%	5.1%	7.0%	4.2%	2.6%
	10	58.5%	18.5%	8.7%	6.6%	6.6%	3.8%	2.8%
	5	60.6%	21.1%	9.5%	7.0%	7.0%	4.0%	2.8%
	2.5	62.3%	25.7%	11.7%	9.4%	7.2%	4.5%	3.4%
45	15	57.0%	13.0%	5.5%	4.4%	9.4%	4.1%	3.2%
	10	56.0%	12.8%	5.9%	4.7%	9.5%	4.0%	3.0%
	5	58.3%	13.0%	5.6%	4.8%	9.3%	3.7%	2.8%
	2.5	60.8%	12.8%	5.5%	4.4%	9.0%	4.1%	2.9%
60	15	52.3%	11.2%	3.9%	2.5%	8.2%	3.4%	2.2%
	10	52.9%	10.4%	3.9%	2.4%	7.1%	2.7%	1.8%
	5	53.7%	9.3%	4.3%	2.5%	7.1%	1.9%	1.2%
	2.5	54.1%	9.6%	2.9%	1.9%	8.1%	1.4%	1.4%
120	15	54.0%	10.6%	4.7%	3.0%	10.4%	4.7%	3.2%
	10	53.6%	10.5%	4.6%	3.1%	10.0%	4.6%	3.4%
	5	55.9%	11.1%	5.2%	3.5%	8.7%	4.5%	3.1%
	2.5	55.9%	10.8%	4.9%	2.9%	11.3%	5.4%	3.4%

Notes: This table presents how is return distributed on the right tail of the statistic λ_t^d, specifically, the percentage of return (on the right tail) that is greater than 0, 1, 1.5, 2 standard deviations of all returns, and also that is smaller than 0, -1, -1.5, -2 standard deviations of all returns.

Table 4.6: S_t^- Evaluations

δ (min)	k(%)	y_t<0	y_t>$\sigma(y_t)$	y_t>1.5$\sigma(y_t)$	y_t>2$\sigma(y_t)$	y_t<-$\sigma(y_t)$	y_t<-1.5$\sigma(y_t)$	y_t<-2$\sigma(y_t)$
5	15	60.2%	5.0%	2.3%	1.4%	17.8%	10.6%	7.8%
	10	62.7%	5.1%	2.5%	1.6%	19.3%	11.5%	8.7%
	5	62.9%	4.9%	2.5%	1.5%	18.7%	11.1%	8.2%
	2.5	63.5%	5.1%	2.7%	1.8%	19.6%	11.5%	8.5%
15	15	59.9%	6.9%	2.8%	2.1%	17.3%	8.4%	6.2%
	10	59.9%	6.2%	2.5%	1.8%	17.7%	8.6%	6.4%
	5	62.3%	6.6%	2.7%	1.6%	19.8%	10.1%	7.4%
	2.5	63.3%	5.4%	2.1%	1.1%	21.2%	10.5%	7.2%
30	15	57.8%	8.8%	4.4%	3.6%	14.7%	7.6%	5.4%
	10	59.2%	7.9%	4.4%	3.8%	16.1%	8.8%	6.2%
	5	62.1%	7.2%	3.6%	2.6%	19.0%	11.3%	7.7%
	2.5	59.0%	9.0%	3.3%	3.3%	19.7%	14.8%	9.8%
45	15	59.1%	5.9%	2.7%	1.6%	9.7%	5.4%	4.8%
	10	56.0%	6.0%	4.8%	2.4%	9.5%	4.8%	3.6%
	5	58.1%	3.2%	3.2%	3.2%	6.5%	0.0%	0.0%
	2.5	61.1%	8.8%	4.3%	2.5%	12.5%	5.8%	3.9%
60	15	52.1%	8.4%	3.7%	2.2%	10.8%	4.3%	2.6%
	10	52.8%	6.8%	2.6%	1.7%	10.1%	4.4%	1.9%
	5	53.4%	8.1%	2.0%	1.2%	8.9%	4.3%	2.5%
	2.5	53.8%	9.0%	1.6%	1.7%	9.7%	3.0%	1.8%
120	15	53.5%	9.9%	4.7%	2.8%	10.2%	4.6%	3.2%
	10	53.9%	9.6%	4.7%	3.3%	10.9%	4.7%	3.5%
	5	56.0%	8.5%	4.1%	3.7%	11.4%	5.0%	3.5%
	2.5	55.4%	10.3%	5.7%	3.3%	10.4%	4.8%	2.9%

Notes: This table presents how is return distributed on the left tail of the statistic λ_t^d, specifically, the percentage of return (on the left tail) that is greater than 0, 1, 1.5, 2 standard deviations of all returns, and also that is smaller than 0, -1, -1.5, -2 standard deviations of all returns.

Table 4.7: Regression Results

δ (min)		5		15		30		45		60		120	
k(%)	Coefficients	Estimates	tStat	Estimates	tStat	Estimates	tStat	Estimates	tStat	Estimates	tStat	Estimates	tStat
	a	0.004	(2.54)	0.003	(1.67)	0.002	(1.07)	0.003	(1.85)	0.003	(1.93)	0.003	(1.92)
15	b_1	0.059	(10.08)	0.039	(6.97)	0.019	(4.79)	0.006	(1.97)	0.001	(0.44)	0.001	(0.23)
	b_2	-0.060	(-10.5)	-0.035	(-7.2)	-0.016	(-3.53)	-0.013	(-2.33)	-0.001	(-0.25)	-0.007	(-1.09)
	a	0.005	(2.73)	0.003	(1.55)	0.002	(1.2)	0.002	(1.61)	0.003	(1.86)	0.003	(1.98)
10	b_1	0.063	(9.76)	0.044	(6.98)	0.025	(5.31)	0.006	(1.96)	0.003	(1.15)	0.001	(0.28)
	b_2	-0.068	(-10.49)	-0.037	(-6.79)	-0.021	(-3.65)	-0.008	(-0.99)	-0.004	(-0.55)	-0.007	(-1.07)
	a	0.004	(2.18)	0.003	(1.85)	0.002	(1.39)	0.002	(1.57)	0.003	(2.05)	0.002	(1.5)
5	b_1	0.066	(8.9)	0.049	(6.74)	0.031	(5.46)	0.009	(2.45)	0.004	(1.15)	0.004	(1.42)
	b_2	-0.067	(-9.48)	-0.044	(-6.6)	-0.028	(-3.71)	-0.001	(-0.17)	-0.001	(-0.13)	-0.007	(-1.03)
	a	0.004	(2.06)	0.003	(1.82)	0.002	(1.41)	0.002	(1.77)	0.003	(2.4)	0.003	(2.15)
2.5	b_1	0.066	(8.54)	0.057	(6.71)	0.040	(5.62)	0.009	(2.51)	0.002	(0.62)	0.001	(0.53)
	b_2	-0.069	(-9.02)	-0.051	(-6.71)	-0.030	(-3)	-0.005	(-1.1)	0.066	(3.5)	-0.007	(-1.28)

Notes: Est. and tStat denote regression coefficient estimates and test statistics. These regression results are based on Newey–West heteroscedasticity and autocorrelation consistent estimator.

Table 4.8: Annualised Mean Return

δ (min)	k%	LS	LS *p.c.*	LO	LO *p.c.*	δ (min)	k%	LS	LS *p.c.*	LO	LO *p.c.*
	15	191.3%	371.6%	141.8%	237.9%		15	109.6%	109.6%	90.5%	90.5%
5	10	185.5%	371.7%	139.6%	243.6%	45	10	99.6%	99.6%	89.3%	89.3%
(144.4%)	5	163.5%	325.9%	127.4%	222.7%	(120.8%)	5	79.2%	79.2%	77.2%	77.2%
	2.5	162.4%	338.3%	126.5%	229.4%		2.5	58.1%	58.1%	58.7%	58.7%
	15	166.5%	230.4%	119.8%	151.3%		15	67.9%	67.9%	65.9%	65.9%
15	10	157.6%	225.4%	120.1%	154.8%	60	10	51.4%	51.4%	56.9%	56.9%
(139.3%)	5	148.8%	219.9%	115.5%	153.9%	(117.6%)	5	57.4%	57.4%	61.0%	61.0%
	2.5	153.1%	230.0%	112.6%	154.9%		2.5	36.6%	36.6%	39.1%	39.1%
	15	128.8%	128.8%	95.5%	95.5%		15	56.6%	56.6%	56.6%	56.6%
30	10	111.5%	111.5%	85.4%	85.4%	120	10	56.8%	56.8%	56.8%	56.8%
(127.8%)	5	87.7%	87.7%	71.7%	71.7%	(116.0%)	5	68.0%	68.0%	68.0%	68.0%
	2.5	73.0%	73.0%	59.2%	59.2%		2.5	47.6%	47.6%	47.6%	47.6%

Notes: This table present annualised mean returns of long-and-short (LS) and long-only (LO) strategy across different k%. *p.c.* denotes the pre-costs strategy. The value in the bracket denotes buy-and-hold returns as a benchmark.

Table 4.9: Sharpe Ratio

δ (min)	k%	LS	LS *p.c.*	LO	LO *p.c.*	δ (min)	k%	LS	LS *p.c.*	LO	LO *p.c.*
	15	3.96**	6.58**	4.36**	6.09**		15	1.68	1.98†	1.58	1.77
5	10	3.73**	6.39**	4.14**	5.75**	45	10	1.82†	2.09*	1.76†	1.98†
(1.31)	5	3.81**	6.14**	3.9**	5.31**	(1.42)	5	1.69	1.98†	1.71	1.98†
	2.5	3.49**	5.77**	3.66**	5.07**		2.5	1.43	1.74	1.47	1.76
	15	3.66**	5.51**	4.16**	5.32**		15	1.42	1.52	1.5	1.57
15	10	3.31**	5.23**	3.95**	5.13**	60	10	1.27	1.36	1.49	1.56
(1.32)	5	3.01**	4.97**	3.57**	4.77**	(1.45)	5	1.48	1.53	1.6	1.64†
	2.5	2.92**	4.76**	3.2**	4.36**		2.5	1.27	1.31	1.38	1.42
	15	1.76†	2.39*	2.47**	3**		15	0.95	1.02	0.95	1.02
30	10	1.82*	2.45**	2.55**	3.1**	120	10	1.01	1.09	1.01	1.09
(1.38)	5	1.74†	2.35**	2.49**	3.04**	(1.49)	5	1.36	1.45	1.36	1.45
	2.5	1.7	2.19*	2.29*	2.76**		2.5	1.1	1.21	1.1	1.21

Notes: This table presents Sharpe ratios of long-and-short (LS) and long-only (LO) strategy across different k%. *p.c.* denotes the pre-costs strategy. The value in the bracket denotes the Sharpe ratio of a buy-and-hold strategy as a benchmark. Ledoit & Wolf (2008)'s test results are also provided with a null hypothesis that the Sharpe ratio of a trading strategy i is greater than that of a buy-and-hold strategy ($H_0 : \widehat{SR}_i - \widehat{SR}_{BaH} > 0$).
** indicates a rejection of the null under 1% significance level,
* indicates a rejection of the null under 5% significance level,
† indicates a rejection of the null under 10% significance level.

Table 4.10: Standard Deviation of Returns and Maximum Drawdown

	$\sigma(y_t)$						MDD					
k%	δ (min)	LS	LO	δ (min)	LS	LO	δ (min)	LS	LO	δ (min)	LS	LO
15		0.27	0.22		0.55	0.50		-8.3%	-5.0%		-32.1%	-33.2%
10	5	0.27	0.22	45	0.47	0.44	5	-7.7%	-5.1%	45	-24.8%	-25.3%
5	(1.10)	0.25	0.22	(0.84)	0.39	0.38	(-43.9%)	-7.2%	-5.1%	(-38.7%)	-18.8%	-19.2%
2.5		0.26	0.23		0.33	0.33		-7.6%	-5.5%		-20.8%	-20.6%
15		0.29	0.21		0.44	0.41		-6.3%	-6.4%		-26.8%	-18.7%
10	15	0.29	0.22	60	0.37	0.36	15	-8.8%	-6.0%	60	-31.5%	-22.7%
5	(1.04)	0.28	0.22	(0.80)	0.37	0.37	(-39.0%)	-9.4%	-5.9%	(-38.6%)	-30.4%	-30.9%
2.5		0.30	0.24		0.27	0.27		-11.2%	-6.2%		-30.0%	-30.0%
15		0.53	0.31		0.55	0.55		-14.7%	-13.6%		-28.1%	-28.1%
10	30	0.45	0.27	120	0.51	0.51	30	-15.4%	-9.3%	120	-22.5%	-22.5%
5	(0.92)	0.36	0.23	(0.77)	0.46	0.46	(-38.8%)	-15.2%	-7.1%	(-38.3%)	-20.3%	-20.3%
2.5		0.33	0.21		0.38	0.38		-15.9%	-5.9%		-34.1%	-34.1%

Notes: This table presents standard deviations and maximum drawdowns of long-and-short (LS) and long-only (LO) strategy returns across different k%. *p.c.* denotes the pre-costs strategy. The value in the bracket denotes the standard deviation and maximum drawdown of a buy-and-hold strategy as a benchmark.

Table 4.11: Calmar Ratio, Omega Ratio and Sortino Ratio

		Calmar Ratio			*Omega Ratio*			*Sortino Ratio*		
δ (min)	k%	LS	LN	BaH	LS	LN	BaH	LS	LN	BaH
5	15	12.59	19.68	2.92	1.57	2.19	1.05	6.25	5.77	1.61
	10	12.79	17.92		1.59	2.19		5.87	5.33	
	5	12.71	16.66		1.74	2.38		6.84	5.93	
	2.5	11.44	15.06		1.75	2.44		6.47	5.72	
15	15	16.36	13.57	2.82	1.65	2.30	1.06	4.24	3.68	1.67
	10	10.36	14.28		1.66	2.42		3.95	3.93	
	5	8.61	13.42		1.71	2.61		3.58	3.86	
	2.5	7.42	11.91		1.88	2.71		3.91	3.52	
30	15	5.91	5.32	2.62	1.28	1.60	1.08	1.37	1.53	1.76
	10	4.92	7.13		1.39	1.83		1.18	1.36	
	5	3.91	7.90		1.50	2.11		0.89	1.22	
	2.5	3.26	7.93		1.66	2.34		0.73	0.96	
45	15	2.40	1.99	2.47	1.30	1.32	1.11	1.18	0.97	1.79
	10	3.03	2.70		1.40	1.42		1.20	1.07	
	5	3.22	3.13		1.45	1.49		1.09	1.08	
	2.5	2.05	2.11		1.46	1.48		0.80	0.81	
60	15	2.04	3.04	2.39	1.26	1.30	1.13	1.17	1.16	1.80
	10	1.27	2.11		1.28	1.36		0.89	1.02	
	5	1.53	1.61		1.46	1.52		0.97	1.03	
	2.5	0.99	1.06		1.50	1.56		0.61	0.67	
120	15	1.61	1.61	2.01	1.19	1.19	1.20	0.90	0.90	1.92
	10	2.07	2.07		1.23	1.23		0.92	0.92	
	5	2.82	2.82		1.41	1.41		1.18	1.18	
	2.5	1.03	1.03		1.37	1.37		0.78	0.78	

Notes: This table shows Calmar ratio (CR), Sortino ratio (SR), and Omega ratio (OR) of long-short and long-only strategy compared to a buy-and-hold strategy. Ratios are calculated as $CR = \frac{AMR}{MDD}, SR = \frac{\overline{y_t^p - r_t}}{\sigma(y_t^p \mathbb{1}_{y_t^p<0})}, OR = \frac{\sum y_t^p \mathbb{1}_{y_t^p>0}}{\sum y_t^p \mathbb{1}_{y_t^p<0}}$

Table 4.12: Annuliased Mean Return and Sharpe Ratio of Trading Strategies with Different Trading Costs

δ (min)	TCs(%)	LS	LO	δ (min)	LS	LO
5	2	102.3%(3.73)	93.9%(4.14)	45	86.4%(1.82)	78.9%(1.76)
	2.5	82.3%(3.04)	82.9%(3.71)		83.1%(1.75)	76.3%(1.7)
	3	62.5%(2.33)	72.1%(3.28)		79.8%(1.68)	73.7%(1.64)
144.4%(1.31)	3.5	43%(1.61)	61.5%(2.83)	120.8%(1.42)	76.5%(1.61)	71.1%(1.59)
	4	23.8%(0.89)	51%(2.38)		73.2%(1.54)	68.5%(1.53)
15	2	95.5%(3.31)	88%(3.95)	60	48%(1.27)	54.3%(1.49)
	2.5	80.5%(2.81)	80.2%(3.64)		47.1%(1.25)	53.7%(1.48)
	3	65.7%(2.31)	72.5%(3.32)		46.2%(1.22)	53%(1.46)
139.3%(1.32)	3.5	51.1%(1.8)	64.9%(3)	117.6%(1.45)	45.4%(1.2)	52.4%(1.44)
	4	36.6%(1.29)	57.4%(2.68)		44.5%(1.18)	51.7%(1.42)
30	2	82%(1.82)	69.3%(2.55)	120	52.8%(1.01)	52.8%(1.01)
	2.5	74.7%(1.66)	65.2%(2.4)		51.8%(0.99)	51.8%(0.99)
	3	67.3%(1.49)	61.2%(2.26)		50.8%(0.98)	50.8%(0.98)
127.8%(1.38)	3.5	60%(1.33)	57.1%(2.12)	116.0%(1.49)	49.8%(0.96)	49.8%(0.96)
	4	52.7%(1.17)	53.1%(1.97)		48.8%(0.94)	48.8%(0.94)

Notes: Sharpe ratios are put in the bracket. Values below δ are for but-and-hold strategies.

Table 4.13: Regression Analysis with Volatility and Return Signs Dummy Variables

		Panel A								Panel B			
δ (min)	k%	c_0		c_1		c_2		c_3		d_0		d_1	
5	15	0.12	(0.39)	1.05	(1.38)	0.66	(1.47)	0.21	(0.48)	0.91	(4.04)	0.36	(1.14)
	10	0.18	(0.58)	0.77	(0.87)	0.55	(1.25)	0.11	(0.24)	0.92	(4.1)	0.21	(0.67)
	5	0.00	(0.01)	0.99	(1.24)	0.58	(1.45)	0.26	(0.66)	0.63	(3.12)	0.62	(2.18)
	2.5	-0.04	(-0.14)	0.94	(1.11)	0.65	(1.55)	0.25	(0.59)	0.58	(2.73)	0.65	(2.17)
15	15	0.55	(1.64)	-0.04	(-0.06)	0.70	(1.49)	-0.35	(-0.75)	0.68	(2.82)	0.75	(2.26)
	10	0.35	(1.07)	0.00	(0.01)	0.72	(1.54)	-0.09	(-0.2)	0.52	(2.18)	0.85	(2.56)
	5	0.17	(0.52)	0.48	(0.7)	0.67	(1.47)	-0.17	(-0.36)	0.61	(2.6)	0.48	(1.48)
	2.5	0.15	(0.44)	0.70	(0.97)	0.58	(1.2)	-0.12	(-0.25)	0.57	(2.32)	0.59	(1.72)
30	15	0.29	(0.47)	0.53	(0.62)	1.06	(1.22)	0.75	(0.87)	0.08	(0.19)	1.67	(2.72)
	10	0.07	(0.14)	1.15	(1.58)	1.05	(1.45)	0.52	(0.71)	0.37	(1)	0.88	(1.7)
	5	0.22	(0.53)	1.01	(1.37)	0.20	(0.33)	0.02	(0.04)	0.15	(0.49)	0.96	(2.29)
	2.5	0.09	(0.25)	0.94	(1.2)	0.30	(0.56)	0.08	(0.16)	0.20	(0.75)	0.70	(1.87)
45	15	0.76	(1.21)	-0.26	(-0.29)	1.07	(1.2)	-0.42	(-0.47)	-1.43	(-3.16)	4.56	(7.26)
	10	0.46	(0.85)	0.47	(0.61)	1.20	(1.56)	-0.35	(-0.46)	-1.27	(-3.28)	4.13	(7.67)
	5	0.36	(0.8)	0.64	(1.01)	0.85	(1.32)	-0.47	(-0.73)	-1.36	(-4.22)	3.94	(8.77)
	2.5	0.14	(0.36)	0.53	(0.99)	0.82	(1.54)	-0.19	(-0.35)	-1.10	(-4.09)	3.05	(8.17)
60	15	0.35	(0.69)	-0.16	(-0.22)	1.16	(1.61)	-0.15	(-0.21)	-2.12	(-5.87)	5.23	(10.51)
	10	0.14	(0.32)	0.50	(0.83)	0.87	(1.43)	-0.21	(-0.35)	-1.60	(-5.27)	3.96	(9.44)
	5	0.07	(0.16)	1.06	(1.77)	0.68	(1.14)	0.03	(0.05)	-1.24	(-4.06)	3.40	(8.11)
	2.5	0.43	(1.39)	-0.20	(-0.44)	0.12	(0.28)	-0.28	(-0.62)	-0.76	(-3.37)	2.11	(6.82)
120	15	0.76	(1.2)	-0.64	(-0.71)	0.16	(0.18)	-0.49	(-0.55)	-2.76	(-6.25)	6.18	(10.21)
	10	0.60	(1.01)	-0.10	(-0.12)	0.27	(0.32)	-0.51	(-0.62)	-2.31	(-5.55)	5.32	(9.35)
	5	0.49	(0.91)	0.43	(0.56)	0.54	(0.71)	-0.48	(-0.64)	-1.62	(-4.26)	4.24	(8.15)
	2.5	0.10	(0.22)	0.36	(0.57)	0.71	(1.14)	0.08	(0.13)	-1.36	(-4.31)	3.37	(7.78)

Notes: Panel A presents the regression results of $y_t^p = c_0 + c_1 Q_1 + c_1 Q_2 + c_1 Q_3 + u_t$, where Q_1 is a dummy variable, which equals 1 when the volatility (v_t) lies in its first quartile. Similarly, Q_2 and Q_3 capture the second and the third quartile of the volatility, respectively. We take quartiles of volatility, which is estimated by Bayesian MCMC, as the regressors. Panel B presents the regression results of $y_t^p = d_0 + d_1 D_t + v_t$, where D_t equals 1 when Bitcoin return (y_t) is positive.

Table 4.14: Regression Analysis with Cryptocurrency Uncertainty Index

		Panel C				Panel D			
δ (min)	k%	e_0		e_1		f_0		f_1	
5	15	4.76	(0.33)	-0.04	(-0.25)	5.77	(0.41)	-0.05	(-0.33)
	10	5.63	(0.41)	-0.04	(-0.33)	6.49	(0.49)	-0.05	(-0.41)
	5	-1.03	(-0.08)	0.02	(0.16)	-0.07	(-0.01)	0.01	(0.09)
	2.5	1.50	(0.11)	-0.01	(-0.04)	2.73	(0.22)	-0.02	(-0.14)
15	15	-1.59	(-1.39)	0.17	(1.49)	-1.55	(-1.41)	0.16	(1.51)
	10	-1.78	(-1.56)	0.19	(1.65)	-1.74	(-1.6)	0.18	(1.69)
	5	-2.23	(-2.1)	0.23	(2.19)	-2.20	(-2.17)	0.23	(2.25)
	2.5	-2.39	(-1.93)	0.24	(2.01)	-2.53	(-2.14)	0.26	(2.22)
30	15	1.88	(0.1)	-0.01	(-0.05)	3.70	(0.2)	-0.03	(-0.15)
	10	-1.53	(-0.1)	0.02	(0.15)	-1.77	(-0.12)	0.03	(0.17)
	5	-5.61	(-0.45)	0.06	(0.5)	-4.69	(-0.39)	0.05	(0.44)
	2.5	-1.16	(-1.04)	0.12	(1.1)	-1.04	(-0.98)	0.11	(1.04)
45	15	-5.28	(-0.26)	0.06	(0.31)	-3.12	(-0.16)	0.04	(0.21)
	10	-1.29	(-0.83)	0.14	(0.89)	-1.10	(-0.74)	0.12	(0.8)
	5	-1.29	(-0.98)	0.13	(1.03)	-1.35	(-1.07)	0.14	(1.12)
	2.5	-1.21	(-1.09)	0.12	(1.13)	-1.22	(-1.15)	0.13	(1.19)
60	15	-2.73	(-0.17)	0.03	(0.21)	-1.70	(-0.11)	0.02	(0.15)
	10	-1.22	(-0.94)	0.13	(0.98)	-1.28	(-1.03)	0.13	(1.07)
	5	-2.27	(-1.63)	0.23	(1.68)	-2.54	(-1.91)	0.26	(1.96)
	2.5	4.40	(0.4)	-0.04	(-0.36)	3.34	(0.32)	-0.03	(-0.28)
120	15	2.27	(1.25)	-0.22	(-1.22)	2.23	(1.29)	-0.21	(-1.25)
	10	1.57	(0.94)	-0.15	(-0.91)	1.72	(1.08)	-0.16	(-1.05)
	5	7.11	(0.45)	-0.06	(-0.41)	10.39	(0.68)	-0.10	(-0.64)
	2.5	-1.25	(-0.11)	0.02	(0.15)	4.23	(0.39)	-0.04	(-0.35)

Notes: This table presents the regression results of the model $y_t^p = e_0 + e_1 CPIPolicy + w_t$ and $y_t^p = f_0 + f_1 CPIPrice + z_t$, where CPIPolicy denotes a cryptocurrency policy uncertainty index and CPIPrice denotes a cryptocurrency price uncertainty index. They are proposed by Lucey et al. (2022). These two indices are available on the website `https://sites.google.com/view/cryptocurrency-indices/home?authuser=0` capture uncertainties caused by policy and regulatory debates and media's attention.

Table 4.15: Priors Specification

Parameters	Prior Distribution
μ	$N(0, 0.1)$
κ	$N(0, 0.1)\mathbb{1}_{\kappa>0}$
θ	$N(0, 0.1)\mathbb{1}_{\theta>0}$
σ_v	$IG(2.5, 0.1)$
ρ	$U(-1, 1)$
ξ^+, ξ^-	$N(0, 5)\mathbb{1}_{\xi^+>0}$
$\boldsymbol{\lambda_0}$	$N(0, 0.03)\mathbb{1}_{\boldsymbol{\lambda_0}>0}$
α	$N(0, 0.3)\mathbb{1}_{\alpha>0}$
β	$N(0, 0.3)\mathbb{1}_{\beta>0}$

Notes: This table presents priors settings in MCMC algorithm of parameter estimations

Chapter 5

Conclusion and Future Research

5.1 Summary and Conclusions

This thesis studies stochastic volatility with jump clustering models and their financial applications. In Chapter 2, we propose a stochastic volatility model embedded with a marked Hawkes process to capture dependency among jumps in different dimensions and the impact from jump sizes. We estimate the model by a Bayesian Markov chain Monte Carlo algorithm. We find jump clustering behaviours in both high-frequency (5-minutes) individual stock and index data. The three dimensions of jumps, namely positive price jumps, negative price jumps and variance jumps impact one another asymmetrically. We also show the importance of intraday periodicity in modelling the dynamics of jump arrivals. In assessing model fitness, we benchmark the model to a range of other models without jump clustering features. We report the deviance information criterion and the Bayes factor, which support the modelling price and variance with jump clustering features. Furthermore, we define clusters of jumps and study their properties. We find a cluster of jumps can cover 2.5 to 6 hours on average. We further conduct a simulation study to investigate the ability of the range of models to reproduce two characteristics in the real data, namely the mean number of jumps in a cluster and the standard deviation of the number of jumps in a cluster. We find the model with jump clustering features

outperforms others in reproducing these characteristics in general.

In Chapter 3, we focus on forecasting using the same model in Chapter 2. In the in-sample period, we estimate the model with 5-minutes individual stock and index data as we do in Chapter 2. In the out-of-sample period, we first fix the static parameters at their posterior mean and use a particle filter to recursively estimate latent variables at each time t, then conduct forward simulations to approximate predictive distributions of variance and return. From these return and variance distributions, we retrieve point forecasts for high-frequency variance, daily realised volatility and two risk measures, namely value-at-risk and expected shortfall. Further, we apply a Kalman filter to correct accumulated forecast errors caused by microstructure noise. Through a simulation study, we show how microstructure noise in high-frequency data can affect parameter estimations and forecasting. We also show the effectiveness of the Kalman filter. We assess our forecasts against a range of benchmark models in many ways, including mean square error, forecast bias, Mincer-Zarnowitz regression, Diebold-Mariano test and conditional performance after different types of jumps. In addition, we also look at 10 and 30 minutes data for comparisons. We have several findings on the forecasting results. Firstly, forecasting results by our model outperform those by other models in general, especially when the forecasting horizon is less than 2 hours. Additionally, forecast performance after negative return jumps is comparatively better than that after other jumps. Including jump clustering effects in forecasting becomes less effective using lower frequency data (10, 30 minutes). Also, our forecasting approach performs better in an individual stock data (Apple Inc.), which is estimated to contain more jumps. Lastly, we backtest forecasts of those two risk measures with two widely used tests. The results also favour our model.

In Chapter 4, we forecast Bitcoin return and its jumps with the same model in the previous two chapters, but a more parsimonious one - a model without intraday periodicity and variance jumps. Because we find that variance jumps exhibit little self/cross-excitation behaviours compared to return jumps. Another reason is to re-

lease a part of computational burden. The forecasting framework is the same as the one in Chapter 2, except that we focus on one-step-ahead prediction on underlying intensities of positive and negative return jumps. We collect Bitcoin tick data and re-sample it with a range of frequencies from 5 minutes to 120 minutes. We evaluate our probabilistic forecasts by a continuous rank probability score. The results favour our model against two other benchmark models. We develop a statistic taking the difference between the predicted probability of positive jumps and that of negative jumps. Tails of this statistic is shown to have strong indications on Bitcoin return's variation. Furthermore, we propose a trading strategy based on the tails of this statistic. We impose restrictions on the strategy in terms of trading costs, trading gaps and liquidity costs. The strategy annualized Sharpe ratio can reach as high as 4.36 including transaction costs. This performance mainly stems from the low standard deviation of strategy returns. The performance of maximum drawdown also confirms the ability of our forecasting framework to mitigate downside risks. We conduct a battery of robustness check, including different trading costs, trading gaps and a range of performance evaluation ratios. We also test whether strategy returns are higher/lower during 1) high/low volatility periods, 2) price increase/decrease periods, and 3) high/low uncertain periods. We find that our strategy works better during the Bitcoin price increases, but it is less relevant to its volatility and uncertainties in the market.

This thesis contributes to current literature in many ways and has several implications for financial practitioners. Firstly, to our knowledge, we are the first to model intraday jump dependency with high-frequency data, and the first to apply the modelling in forecasting on various variables, in which market practitioners are interested. Our research provides financial market participants a guide on the likelihood of further jumps when they observe a jump happening. We also provide frameworks to incorporate this jump dependency in forecasting risks, return and return jumps. It is also valuable for the field of risk management.

5.2 Future Research

In this section, we discuss questions left in this thesis and some opportunities for further research.

Option Pricing In Chapter 2, we propose a continuous-time asset pricing model. We assess the fitness that stock and index data lend to the model in many aspects. One aspect we did not consider is its calibration to option data. This is a typical application of continuous-time stochastic volatility models in the literature (see e.g. Heston 1993, Bates 1996, Duffie et al. 2000, Eraker 2004). However, calibrating the model proposed in Chapter 2 to option data may also be subject to several difficulties, which are also reasons for us not considering option pricing at the current stage. Specifically, the model we propose is for high-frequency data, while the availability of high-frequency data is an obstacle of this practice. Additionally, it is also important to consider the computational burden in practice. We use a hybrid of parametric and non-parametric estimation methods in Chapter 2 to reduce the burden caused by high-frequency data. Additionally, the estimation is under a physical measure. Whether they are appropriate in a risk-neutral measure awaits further validation. Alternatively, we may need a new model estimation technique to guarantee the consistency of estimated parameters within acceptable computational times. However, the high-frequency option pricing subject remains blank in the literature. The intraday dynamics of options can be valuable to the financial market.

Jump Predictions in Various Markets In Chapter 4, we forecast Bitcoin return and its jumps. The intuition we study on Bitcoin is that the Bitcoin market is highly inefficient, contains a large number of jumps and, hence, has a high potential to exhibit jump self/cross-excitation. In this sense, the model can be applied to any assets that have the potential to exhibit jump self/cross-excitation. A number of studies show the self-excitation exhibited in foreign exchange markets (see e.g. Lee 2012, Lee & Wang 2020). However, their prediction approach mainly relies on

macro-information releases, which greatly differs from the forecasting framework in Chapter 3. In addition to the foreign exchange markets, the commodity market is also worth an investigation. Diewald et al. (2015), Nguyen & Prokopczuk (2019)'s studies give an overview of jumps in the commodity market. They show a strong inter-correlation in the commodity, to which our model can perfectly apply.

Asset Pricing In Chapter 2, we only apply the jump clustering model to 5 individual stocks and an index data. We show the different extent of jump clustering among these assets. A following natural question would be what is the implication of different extent of jump clustering in different assets. More importantly, is the extent of jump clustering an important factor and is it priced in the cross-sectional stock data? In Chapter 2, we provide an explanation for jump clustering, which is also our intuition for conducting this research. Investors in the market are not homogeneous regarding how they are informed. After the releases of a company's information, the most informed investor will trade according to the information, which creates a jump in stock prices, which follows other jumps created by less-informed investors. In other words, if investors of a company are equally well-informed, there less-likely exists jump clustering in the company's stock prices. Therefore, the extent of jump clustering of an asset's prices may have implications on how the asset's investors are informed and the efficiency of the asset.

References

Aït-Sahalia, Y., Cacho-Diaz, J. & Laeven, R. J. (2015), 'Modeling financial contagion using mutually exciting jump processes', *Journal of Financial Economics* **117**(3), 585–606.

Andersen, T. G. & Bollerslev, T. (1997), 'Intraday periodicity and volatility persistence in financial markets', *Journal of Empirical Finance* **4**(2-3), 115–158.

Andersen, T. G., Bollerslev, T. & Diebold, F. X. (2007), 'Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility', *The Review of Economics and Statistics* **89**(4), 701–720.

Andersen, T. G., Thyrsgaard, M. & Todorov, V. (2019), 'Time-varying periodicity in intraday volatility', *Journal of the American Statistical Association* **114**(2), 1–26.

Anyfantaki, S., Arvanitis, S. & Topaloglou, N. (2021), 'Diversification benefits in the cryptocurrency market under mild explosivity', *European Journal of Operational Research* **295**(1), 378–393.

Asgharian, H. & Bengtsson, C. (2006), 'Jump spillover in international equity markets', *Journal of Financial Econometrics* **4**(2), 167–203.

Atsalakis, G. S., Atsalaki, I. G., Pasiouras, F. & Zopounidis, C. (2019), 'Bitcoin price forecasting with neuro-fuzzy techniques', *European Journal of Operational Research* **276**(2), 770–780.

Bakshi, G., Cao, C. & Chen, Z. (1997), 'Empirical performance of alternative option pricing models', *The Journal of Finance* **52**(5), 2003–2049.

Bandi, F. M. & Reno, R. (2016), 'Price and volatility co-jumps', *Journal of Financial Economics* **119**(1), 107–146.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2008), 'Designing realized kernels to measure the ex post variation of equity prices in the presence of noise', *Econometrica* **76**(6), 1481–1536.

Barndorff-Nielsen, O. E. & Shephard, N. (2006), 'Impact of jumps on returns and realised variances: econometric analysis of time-deformed Lévy processes', *Journal of Econometrics* **131**(1-2), 217–252.

Bates, D. S. (1996), 'Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options', *The Review of Financial Studies* **9**(1), 69–107.

Black, F. (1976), 'Studies of stock market volatility changes', *1976 Proceedings of the American statistical association bisiness and economic statistics section* .

Boswijk, H. P., Laeven, R. J. & Yang, X. (2018), 'Testing for self-excitation in jumps', *Journal of Econometrics* **203**(2), 256–266.

Boudt, K., Croux, C. & Laurent, S. (2011), 'Robust estimation of intraweek periodicity in volatility and jump detection', *Journal of Empirical Finance* **18**(2), 353–367.

Brandvold, M., Molnár, P., Vagstad, K. & Valstad, O. C. A. (2015), 'Price discovery on Bitcoin exchanges', *Journal of International Financial Markets, Institutions and Money* **36**, 18–35.

Brauneis, A., Mestel, R., Riordan, R. & Theissen, E. (2021), 'How to measure the liquidity of cryptocurrency markets?', *Journal of Banking & Finance* **124**, 106041.

Brier, G. W. et al. (1950), 'Verification of forecasts expressed in terms of probability', *Monthly weather review* **78**(1), 1–3.

Brillinger, D. R. (1988), 'Some statistical methods for random process data from seismology and neurophysiology', *The Annals of Statistics* **16**(1), 1–54.

Broadie, M., Chernov, M. & Johannes, M. (2007), 'Model specification and risk premia: Evidence from futures options', *The Journal of Finance* **62**(3), 1453–1490.

Carr, P. & Wu, L. (2003), 'The finite moment log stable process and option pricing', *The Journal of Finance* **58**(2), 753–777.

Chaim, P. & Laurini, M. P. (2018), 'Volatility and return jumps in Bitcoin', *Economics Letters* **173**, 158–163.

Chen, J., Urquhart, A. & Clements, M. P. (2021), 'Modelling price and variance jump clustering using the marked Hawkes process', *Available at SSRN 3992885* . **URL:** *http://dx.doi.org/10.2139/ssrn.3992885*

Chib, S. (1995), 'Marginal likelihood from the gibbs output', *Journal of the American Statistical Association* **90**(432), 1313–1321.

Chib, S. & Jeliazkov, I. (2001), 'Marginal likelihood from the metropolis–hastings output', *Journal of the American Statistical Association* **96**(453), 270–281.

Clements, M. P. & Krolzig, H.-M. (2004), 'Can regime-switching models reproduce the business cycle features of us aggregate consumption, investment and output?', *International Journal of Finance & Economics* **9**(1), 1–14.

Clinet, S. & Potiron, Y. (2021), 'Disentangling sources of high frequency market microstructure noise', *Journal of Business & Economic Statistics* **39**(1), 18–39.

Corbet, S., Lucey, B., Urquhart, A. & Yarovaya, L. (2019), 'Cryptocurrencies as a financial asset: A systematic analysis', *International Review of Financial Analysis* **62**, 182–199.

Corsi, F. (2009), 'A simple approximate long-memory model of realized volatility', *Journal of Financial Econometrics* **7**(2), 174–196.

Creal, D. (2012), 'A survey of sequential monte carlo methods for economics and finance', *Econometric Reviews* **31**(3), 245–296.

Daley, D. J., Vere-Jones, D. et al. (2003), 'An introduction to the theory of point processes, volume 1: Elementary theory and methods', *Verlag New York Berlin Heidelberg: Springer* **11**, 69.

Diebold, F. X. & Mariano, R. S. (1995), 'Comparing predictive accuracy', *Journal of Business & Economic Statistics* **20**(1), 134–144.

Diebold, F. X. & Mariano, R. S. (2002), 'Comparing predictive accuracy', *Journal of Business & Economic Statistics* **20**(1), 134–144.

Diewald, L., Prokopczuk, M. & Simen, C. W. (2015), 'Time-variations in commodity price jumps', *Journal of Empirical Finance* **31**, 72–84.

Duffie, D., Pan, J. & Singleton, K. (2000), 'Transform analysis and asset pricing for affine jump-diffusions', *Econometrica* **68**(6), 1343–1376.

Engle, R. F. & Manganelli, S. (2004), 'Caviar: Conditional autoregressive value at risk by regression quantiles', *Journal of Business & Economic Statistics* **22**(4), 367–381.

Eraker, B. (2004), 'Do stock prices and volatility jump? reconciling evidence from spot and option prices', *The Journal of Finance* **59**(3), 1367–1403.

Evensen, G. (1994), 'Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics', *Journal of Geophysical Research: Oceans* **99**(C5), 10143–10162.

Figueroa-López, J. E. & Mancini, C. (2019), 'Optimum thresholding using mean and conditional mean squared error', *Journal of Econometrics* **208**(1), 179–210.

Fissler, T. & Ziegel, J. F. (2016), 'Higher order elicitability and Osband's principle', *The Annals of Statistics* **44**(4), 1680–1707.

Foley, S., Karlsen, J. R. & Putniņš, T. J. (2019), 'Sex, drugs, and Bitcoin: How much illegal activity is financed through cryptocurrencies?', *The Review of Financial Studies* **32**(5), 1798–1853.

Foschi, R., Lilla, F. & Mancini, C. (2019), 'Warnings about future jumps: properties of the exponential Hawkes model', *Available at SSRN 3459443* .
URL: *http://dx.doi.org/10.2139/ssrn.3459443*

Fulop, A., Li, J. & Yu, J. (2015), 'Self-exciting jumps, learning, and asset pricing implications', *The Review of Financial Studies* **28**(3), 876–912.

Gandal, N., Hamrick, J., Moore, T. & Oberman, T. (2018), 'Price manipulation in the Bitcoin ecosystem', *Journal of Monetary Economics* **95**, 86–96.

Gneiting, T., Balabdaoui, F. & Raftery, A. E. (2007), 'Probabilistic forecasts, calibration and sharpness', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**(2), 243–268.

Gneiting, T. & Raftery, A. E. (2007), 'Strictly proper scoring rules, prediction, and estimation', *Journal of the American statistical Association* **102**(477), 359–378.

Gresnigt, F., Kole, E. & Franses, P. H. (2016), 'Specification testing in Hawkes models', *Journal of Financial Econometrics* **15**(1), 139–171.

Gronwald, M. (2019), 'Is Bitcoin a commodity? on price jumps, demand shocks, and certainty of supply', *Journal of International Money and Finance* **97**, 86–92.

Grossman, S. (1976), 'On the efficiency of competitive stock markets where trades have diverse information', *The Journal of Finance* **31**(2), 573–585.

Guesmi, K., Saadi, S., Abid, I. & Ftiti, Z. (2019), 'Portfolio diversification with virtual currency: Evidence from Bitcoin', *International Review of Financial Analysis* **63**, 431–437.

Hansen, P. R. & Huang, Z. (2016), 'Exponential GARCH modeling with realized measures of volatility', *Journal of Business & Economic Statistics* **34**(2), 269–287.

Hansen, P. R., Huang, Z. & Shek, H. H. (2012), 'Realized GARCH: a joint model for returns and realized measures of volatility', *Journal of Applied Econometrics* **27**(6), 877–906.

Hawkes, A. G. (1971*a*), 'Point spectra of some mutually exciting point processes', *Journal of the Royal Statistical Society: Series B (Methodological)* **33**(3), 438–443.

Hawkes, A. G. (1971*b*), 'Spectra of some self-exciting and mutually exciting point processes', *Biometrika* **58**(1), 83–90.

Hecq, A., Laurent, S. & Palm, F. C. (2012), 'Common intraday periodicity', *Journal of Financial Econometrics* **10**(2), 325–353.

Hersbach, H. (2000), 'Decomposition of the continuous ranked probability score for ensemble prediction systems', *Weather and Forecasting* **15**(5), 559–570.

Hess, G. D. & Iwata, S. (1997), 'Measuring and comparing business-cycle features', *Journal of Business & Economic Statistics* **15**(4), 432–444.

Heston, S. L. (1993), 'A closed-form solution for options with stochastic volatility with applications to bond and currency options', *The Review of Financial Studies* **6**(2), 327–343.

Jacod, J., Li, Y. & Zheng, X. (2017), 'Statistical properties of microstructure noise', *Econometrica* **85**(4), 1133–1174.

Jacod, J. & Todorov, V. (2009), 'Testing for common arrivals of jumps for discretely observed multidimensional processes', *The Annals of Statistics* **37**(4), 1792–1838.

Jacod, J. & Todorov, V. (2010), 'Do price and volatility jump together?', *The Annals of Applied Probability* **20**(4), 1425–1469.

Jacquier, E., Polson, N. G. & Rossi, P. E. (2004), 'Bayesian analysis of stochastic volatility models with fat-tails and correlated errors', *Journal of Econometrics* **122**(1), 185–212.

Jiang, S., Li, Y., Wang, S. & Zhao, L. (2022), 'Blockchain competition: The trade-off between platform stability and efficiency', *European Journal of Operational Research* **296**(3), 1084–1097.

Kass, R. E. & Raftery, A. E. (1995), 'Bayes factors', *Journal of the American Statistical Association* **90**(430), 773–795.

Ketelbuters, J.-J. & Hainaut, D. (2022), 'CDS pricing with fractional Hawkes processes', *European Journal of Operational Research* **297**(3), 1139–1150.

Klein, T., Thu, H. P. & Walther, T. (2018), 'Bitcoin is not the new gold–a comparison of volatility, correlation, and portfolio performance', *International Review of Financial Analysis* **59**, 105–116.

Lazar, E. & Qi, S. (2022), 'Model risk in the over-the-counter market', *European Journal of Operational Research* **298**(2), 769–784.

Ledoit, O. & Wolf, M. (2008), 'Robust performance hypothesis testing with the sharpe ratio', *Journal of Empirical Finance* **15**(5), 850–859.

Lee, K. & Seo, B. K. (2017), 'Marked Hawkes process modeling of price dynamics and volatility estimation', *Journal of Empirical Finance* **40**, 174–200.

Lee, K. & Seo, B. K. (2022), 'Modeling bid and ask price dynamics with an extended Hawkes process and its empirical applications for high-frequency stock market data', *Journal of Financial Econometrics* **1**(1), 1–14.

Lee, S. S. (2012), 'Jumps and information flow in financial markets', *The Review of Financial Studies* **25**(2), 439–479.

Lee, S. S. & Wang, M. (2019), 'The impact of jumps on carry trade returns', *Journal of Financial Economics* **131**(2), 433–455.

Lee, S. S. & Wang, M. (2020), 'Tales of tails: Jumps in currency markets', *Journal of Financial Markets* **48**, 100497.

Li, J. & Zinna, G. (2018), 'The variance risk premium: Components, term structures, and stock return predictability', *Journal of Business & Economic Statistics* **36**(3), 411–425.

Li, Y., Nolte, I., Yu, S. et al. (2021), 'Testing for jumps in a discretely observed price process with endogenous sampling times', *Sandra and Yu, Shifan, Testing for Jumps in a Discretely Observed Price Process with Endogenous Sampling Times (November 12, 2021)* .

Liniger, T. J. (2009), Multivariate Hawkes processes, PhD thesis, ETH Zurich.

Liu, G., Jin, Z. & Li, S. (2021), 'Household lifetime strategies under a self-contagious market', *European Journal of Operational Research* **288**(3), 935–952.

Lucey, B. M., Vigne, S. A., Yarovaya, L. & Wang, Y. (2022), 'The cryptocurrency uncertainty index', *Finance Research Letters* **45**, 102147.

Mancini, C., Mattiussi, V. & Renò, R. (2015), 'Spot volatility estimation using delta sequences', *Finance and Stochastics* **19**(2), 261–293.

Maneesoonthorn, W., Forbes, C. S. & Martin, G. M. (2017), 'Inference on self-exciting jumps in prices and volatility using high-frequency measures', *Journal of Applied Econometrics* **32**(3), 504–532.

Maneesoonthorn, W., Martin, G. M. & Forbes, C. S. (2020), 'High-frequency jump tests: Which test should we use?', *Journal of Econometrics* **219**(2), 478–487.

Matheson, J. E. & Winkler, R. L. (1976), 'Scoring rules for continuous probability distributions', *Management science* **22**(10), 1087–1096.

Merton, R. C. (1976), 'Option pricing when underlying stock returns are discontinuous', *Journal of Financial Economics* **3**(1-2), 125–144.

Mincer, J. A. & Zarnowitz, V. (1969), The evaluation of economic forecasts, *in* 'Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance', NBER, pp. 3–46.

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P. & Tita, G. E. (2011), 'Self-exciting point process modeling of crime', *Journal of the American Statistical Association* **106**(493), 100–108.

Moskowitz, T. J., Ooi, Y. H. & Pedersen, L. H. (2012), 'Time series momentum', *Journal of Financial Economics* **104**(2), 228–250.

Nadarajah, S. & Chu, J. (2017), 'On the inefficiency of Bitcoin', *Economics Letters* **150**, 6–9.

Nguyen, D. B. B. & Prokopczuk, M. (2019), 'Jumps in commodity markets', *Journal of Commodity Markets* **13**, 55–70.

Novotnỳ, J., Petrov, D. & Urga, G. (2015), 'Trading price jump clusters in foreign exchange markets', *Journal of Financial Markets* **24**, 66–92.

Pan, J. (2002), 'The jump-risk premia implicit in options: Evidence from an integrated time-series study', *Journal of Financial Economics* **63**(1), 3–50.

Patton, A. J. & Sheppard, K. (2015), 'Good volatility, bad volatility: Signed jumps and the persistence of volatility', *Review of Economics and Statistics* **97**(3), 683–697.

Patton, A. J., Ziegel, J. F. & Chen, R. (2019), 'Dynamic semiparametric models for expected shortfall (and value-at-risk)', *Journal of Econometrics* **211**(2), 388–413.

Pitt, M. K. & Shephard, N. (1999), 'Filtering via simulation: Auxiliary particle filters', *Journal of the American Statistical Association* **94**(446), 590–599.

Rasmussen, J. G. (2013), 'Bayesian inference for Hawkes processes', *Methodology and Computing in Applied Probability* **15**(3), 623–642.

Reynaud-Bouret, P. & Schbath, S. (2010), 'Adaptive estimation for Hawkes processes; application to genome analysis', *The Annals of Statistics* **38**(5), 2781–2822.

Rousseeuw, P. J. & Leroy, A. M. (1988), 'A robust scale estimator based on the shortest half', *Statistica Neerlandica* **42**(2), 103–116.

Sapkota, N. (2022), 'News-based sentiment and Bitcoin volatility', *International Review of Financial Analysis* **82**, 102183.
URL: *https://www.sciencedirect.com/science/article/pii/S1057521922001454*

Schnaubelt, M. (2022), 'Deep reinforcement learning for the optimal placement of cryptocurrency limit orders', *European Journal of Operational Research* **296**(3), 993–1006.

Shen, D., Urquhart, A. & Wang, P. (2020), 'Forecasting the volatility of Bitcoin: The importance of jumps and structural breaks', *European Financial Management* **26**(5), 1294–1323.

Shumway, R. H. & Stoffer, D. S. (1982), 'An approach to time series smoothing and forecasting using the em algorithm', *Journal of Time Series Analysis* **3**(4), 253–264.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Van Der Linde, A. (2002), 'Bayesian measures of model complexity and fit', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**(4), 583–639.

Stroud, J. R. & Johannes, M. S. (2014), 'Bayesian modeling and forecasting of 24-hour high-frequency volatility', *Journal of the American Statistical Association* **109**(508), 1368–1384.

Todorov, V. & Tauchen, G. (2011), 'Volatility jumps', *Journal of Business & Economic Statistics* **29**(3), 356–371.

Urquhart, A. (2016), 'The inefficiency of Bitcoin', *Economics Letters* **148**, 80–82.

Urquhart, A. (2018), 'What causes the attention of Bitcoin?', *Economics Letters* **166**, 40–44.

Wei, W. C. (2018), 'Liquidity and market efficiency in cryptocurrencies', *Economics Letters* **168**, 21–24.

Zhang, Z., Ren, D., Lan, Y. & Yang, S. (2022), 'Price competition and blockchain adoption in retailing markets', *European Journal of Operational Research* **300**(2), 647–660.